

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:03:31 ; Search time 63 Seconds  
(without alignments)  
1015.243 Million cell updates/sec

Title: US-09-830-980-1  
Perfect score: 2574  
Sequence: 1 MQETDTEQATPTIQARLV.....PDGSRVASGGKDKVIKLWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2558.5	99.4	479	21	AAAY79678
2	2554.5	99.2	481	22	ABBS59486
3	1521	59.1	485	22	ABBS68284
4	1520	59.1	485	22	ABBS92844
5	1520	59.1	485	22	ABBS68516
6	1520	59.1	485	23	ABBS97306
7	1452	56.4	484	22	ABBS68282
8	1335.5	51.9	471	21	AAG23141
9	1335.5	51.9	471	21	AAG46521
10	1335.5	51.9	473	21	AAG23140
					Drosophila Notchle
					Drosophila melanog
					Amino acid sequenc
					Human protein sequ
					Human GTP-binding
					Novel human protei
					Amino acid sequenc
					Arabidopsis thalia
					Arabidopsis thalia

11	1335.5	51.9	473	21	AAG46520	Arabidopsis thalia
12	1335.5	51.9	490	21	AAG23139	Arabidopsis thalia
13	1335.5	51.9	490	21	AAG46519	Arabidopsis thalia
14	1058.5	41.1	514	16	AAR85881	WD-40 domain-contg
15	819	31.8	261	22	AAB68283	Amino acid sequenc
16	419	16.3	317	21	AAB63186	Human secreted pro
17	419	16.3	334	22	AAB68529	Human GTP-binding
18	419	16.3	334	23	ABBS7345	Novel human protei
19	415	16.1	361	22	ABBS68576	Drosophila melanog
20	412.5	16.0	700	22	ABBS6376	Drosophila melanog
21	399	15.5	330	22	AAB93659	Human protein sequ
22	398.5	15.5	341	21	AAG38744	Arabidopsis thalia
23	389	15.1	411	22	ABBS62260	Drosophila melanog
24	372.5	14.5	514	22	AAB95225	Human protein sequ
25	372	14.5	323	21	AAB63185	Gene 3 human secre
26	371	14.4	542	23	ABP41760	Human ovarian anti
27	355	13.8	333	21	AAG05554	Arabidopsis thalia
28	354	13.8	409	16	AAR70002	OPDE 45 kDa subuni
29	354	13.8	410	16	AAR70005	OPDE 45 kDa subuni
30	348	13.5	478	22	AAM93784	Human polypeptide,
31	348	13.5	521	22	ABBI0141	Human cDNA SEQ ID
32	344	13.4	159	20	AAI10919	Amino acid sequenc
33	343	13.3	407	22	AAM93675	Human polypeptide,
34	340.5	13.2	409	16	AAR85868	WD-40 domain-contg
35	335	13.0	358	22	ABBS5223	Drosophila melanog
36	331	12.9	584	22	ABG21351	Novel human diagno
37	330	12.8	423	22	AAB94261	Human protein sequ
38	330	12.8	423	22	AAB94262	Human protein sequ
39	327	12.7	388	21	AAB43313	Human ORFX ORF3077
40	321	12.5	423	22	AAB93290	Human protein sequ
41	320	12.4	423	20	AAI22499	Human secreted pro
42	320	12.4	423	22	AAU39068	Human secreted pro
43	320	12.4	423	22	AAM39569	Human polypeptide
44	320	12.4	423	23	ABBS5777	Human polypeptide
45	320	12.4	447	22	AAM41355	Human polypeptide

ALIGNMENTS

RESULT 1

AAAY79678

ID AAAY79678 standard; Protein; 479 AA.

XX AAAY79678;

AC AAAY79678;

DT 29-AUG-2000 (first entry)

XX Drosophila Notchless protein.

XX Notchless; Nle gene; Notch; signalling; neurodegenerative disease;

KW cancer; diagnosis; cytostatic; neuroprotective; therapy.

XX Drosophila melanogaster.

XX Key Location/Qualifiers

FT Domain 27..105

FT Misc-difference 67 /note= "Nle domain"

FT /note= "sequence deduced from nucleotide sequence  
has an additional Lys residue between  
Lys-67 and Ser-68"

FT Misc-difference 282 /note= "encoded by CAA"

XX WO200026364-A1.

XX 11-MAY-2000.

XX 03-NOV-1999; 99WO-IB01891.

XX 03-NOV-1998; 98GB-0024045.

XX

PA (EUMO-) EURO MOLECULAR BIOLOGY LAB.  
XX Cohen S, Bouwmeester A, Royet J;  
XX WPI; 2000-365613/31.  
DR N-PSDB; AAA27739.

XX Novel Notchless protein and nucleic acids encoding them useful for  
PT treating and preventing cancer and neurodegenerative diseases -  
XX

PS Claim 1; Page 44; 52pp; English.

XX The present sequence is that of Notchless, a novel protein of  
CC Drosophila. Notchless was identified in a screen for dominant  
CC modifiers of a Notch mutant phenotype in the Drosophila wing. The  
CC mutant dominantly suppressed the wing notching phenotype of  
CC notchoid mutations, and the Notchless protein was shown to bind to  
CC the cytoplasmic domain of Notch. Notchless modified Notch  
CC signalling activity in a variety of Notch-dependent signalling  
CC process in both Drosophila and Xenopus embryos. The Notchless  
CC protein has a novel highly conserved N-terminal domain followed by  
CC 9 WD40 repeats. Notchless, and nucleic acids encoding it, can  
CC be used in methods for the diagnosis and therapy of certain diseases,  
CC particularly cancer and neurodegenerative diseases (claimed). A  
CC Notchless mutant in a sensitised Notch genetic background is used  
CC in a claimed method for identifying compounds capable of modifying  
CC the levels of expression or activity of a Notch protein.

XX Sequence 479 AA;

Query Match 99.4%; Score 2558.5; DB 21; Length 479;  
Best Local Similarity 99.8%; Pred. No. 1.1e-232;  
Matches 479; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MQETDTEQATPHTIQARLVYTGEGAGPPIDLPAGITTTQQLGLICNALLKNEEATPYLFF 60  
DB 1 MQETDTEQATPHTIQARLVYTGEGAGPPIDLPAGITTTQQLGLICNALLKNEEATPYLFF 60  
QY 61 VGEDEIKKSLEDTLDLASVDTENVIDIVYQPAVFKVRPVTRCTSSMPGHAEAVVSLNFS 120  
DB 61 VGEDEI-KSLEDTLDLASVDTENVIDIVYQPAVFKVRPVTRCTSSMPGHAEAVVSLNFS 119  
QY 121 PDGAHLASGSGDTTVRLWDLNLTETPHFTCTGKQWVLCVSWAPDGKRLASGCKAGSIILW 180  
DB 120 PDGAHLASGSGDTTVRLWDLNLTETPHFTCTGKQWVLCVSWAPDGKRLASGCKAGSIILW 179  
QY 181 DPETGQKGRPLSGHKKHINCLAWEPYHRDPECKRLASASGDCDCRIWDVKLGQCLMNI 240  
DB 180 DPETGQKGRPLSGHKKHINCLAWEPYHRDPECKRLASASGDCDCRIWDVKLGQCLMNI 239  
QY 241 GHTNAVAVRWGGAGLIYTSKDRTVKMWRAADGILCRTFSGHAHWVNNIALSTDYVLR 300  
DB 240 GHTNAVAVRWGGAGLIYTSKDRTVKMWRAADGILCRTFSGHAHWVNNIALSTDYVLR 299  
QY 301 GPFHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDDNTLYLWRNNQKCV 360  
DB 300 GPFHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDDNTLYLWRNNQKCV 359  
QY 361 ERMTHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMATFRGHVQAVYTVAWSAD 420  
DB 360 ERMTHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMATFRGHVQAVYTVAWSAD 419  
QY 421 SRLIVSGSKDSTLKVWSVQTKKLAQELPGHADEVFQVDWAPDGSRVASGGKDKVIKLWAY 480  
DB 420 SRLIVSGSKDSTLKVWSVQTKKLAQELPGHADEVFQVDWAPDGSRVASGGKDKVIKLWAY 479

RESULT 2  
ABB59486  
ID ABB59486 standard; Protein; 481 AA.  
XX  
AC ABB59486;  
XX

DT 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 5250.  
DE  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL03589.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 5250; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 481 AA;

Query Match 99.2%; Score 2554.5; DB 22; Length 481;  
Best Local Similarity 99.6%; Pred. No. 2.7e-232;  
Matches 479; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 MQETDTEQATPHTIQARLVY-TGEEAGPPIDLPAGITTTQQLGLICNALLKNEEATPYLF 59  
DB 1 MQETDTEQATPHTIQARLVSDTGEAGPPIDLPAGITTTQQLGLICNALLKNEEATPYLF 60  
QY 60 FVGEDEIKKSLEDTLDLASVDTENVIDIVYQPAVFKVRPVTRCTSSMPGHAEAVVSLNF 119  
DB 61 FVGEDEIKKSLEDTLDLASVDTENVIDIVYQPAVFKVRPVTRCTSSMPGHAEAVVSLNF 120  
QY 120 SPDGAHLASGSGDTTVRLWDLNLTETPHFTCTGKQWVLCVSWAPDGKRLASGCKAGSIIL 179  
DB 121 SPDGAHLASGSGDTTVRLWDLNLTETPHFTCTGKQWVLCVSWAPDGKRLASGCKAGSIIL 180  
QY 180 WDPETGQKGRPLSGHKKHINCLAWEPYHRDPECKRLASASGDCDCRIWDVKLGQCLMNI 239  
DB 181 WDPETGQKGRPLSGHKKHINCLAWEPYHRDPECKRLASASGDCDCRIWDVKLGQCLMNI 240  
QY 240 AGHTNAVAVRWGGAGLIYTSKDRTVKMWRAADGILCRTFSGHAHWVNNIALSTDYVLR 299  
DB 241 AGHTNAVAVRWGGAGLIYTSKDRTVKMWRAADGILCRTFSGHAHWVNNIALSTDYVLR 300  
QY 300 TGPFPVVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDDNTLYLWRNNQKVC 359  
DB 301 TGPFPVVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDDNTLYLWRNNQKVC 360

QY 360 VERMTGHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDQGYMATFRGHVQAVYTVAWSA 419  
Db 361 VERMTGHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDQGYMATFRGHVQAVYTVAWSA 420  
QY 420 DSRLIVSGSKDSTLKVMSVQTKKLAQELPGHADEVFGVDWAPDGSRSVSGGKDKVIKWA 479  
Db 421 DSRLIVSGSKDSTLKVMSVQTKKLAQELPGHADEVFGVDWAPDGSRSVSGGKDKVIKWA 480  
QY 480 Y 480  
Db 481 Y 481

RESULT 3  
AAB68284  
ID AAB68284 standard; Protein; 485 AA.  
XX  
AC AAB68284;  
XX  
DT 09-JUL-2001 (first entry)  
XX  
DE Amino acid sequence of a human Tzap gene cDNA clone.  
XX  
KW Tzap; T cell activation; immune response; transplant rejection;  
KW bone marrow transplantation; rheumatoid arthritis; lupus erythematosus;  
KW multiple sclerosis; encephalitis; vasculitis; diabetes mellitus;  
KW pancreatitis; gastritis; thyroiditis; malignant disorder; asthma;  
KW lepramatosus; gastritis; skin tumour; adrenal tumour; lung tumour;  
KW wound healing; growth disorder; inflammatory disease; infectious disease.  
XX

OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 379 /note= "unspecified residue encoded by NTC"  
FT  
XX  
PN WO200132614-A2.  
XX  
PD 10-MAY-2001.  
XX  
PF 30-OCT-2000; 2000WO-EP10670.  
XX  
PR 01-NOV-1999; 99US-0162675.  
PR 25-FEB-2000; 2000US-0185016.  
XX  
PA (UTKU/) UTKU N.  
XX  
XX Utku N;  
PI  
DR WPI; 2001-316410/33.  
DR N-PSDB; AAF85075.

XX  
PT Novel polynucleotide encoding Tzap protein involved in T cell  
PT activation, useful for diagnosing and treating diseases involving T  
PT cell activation, for treating organ transplantation rejection,  
PT rheumatoid arthritis  
XX  
PS Example; Page 56-57; 68pp; English.  
XX  
CC The present sequence represents a Tzap protein. The Tzap gene is  
CC involved in T cell activation, and in the modulation of immune  
CC responses. Tzap polynucleotides and polypeptides are useful for  
CC diagnosing or treating acute and chronic diseases involving T cell  
CC activation and Th1 and Th2 immune response, for the treatment of acute  
CC and chronic rejection of allo- and xeno organ transplants and bone  
CC marrow transplantation, for the treatment of rheumatoid arthritis, lupus  
CC erythematosus, multiple sclerosis, encephalitis, vasculitis, diabetes  
CC mellitus, pancreatitis, gastritis, thyroiditis, for the treatment of  
CC malignant disorders of T, B or NK cells, for the treatment of asthma,  
CC lepramatosus, Helicobacter pylori associated gastritis or for the  
CC treatment of skin tumours, adrenal tumours or lung tumours, wound  
CC healing, growth disorders inflammatory and/or infectious diseases.  
XX

SQ Sequence 485 AA;  
Query Match 59.1%; Score 1521; DB 22; Length 485;  
Best Local Similarity 58.6%; Pred. No. 1.2e-134;  
Matches 279; Conservative 73; Mismatches 118; Indels 6; Gaps 3;  
QY 9 EATPHTIOARLYVTGEEA---GPPIDLPAGITTTQQLGLICNALKNEEATPYLFFVGED 64  
Db 8 EAVARDVQRLLVQFQDEGGQLLGSPEDVPDITPDRILQLVNCALLAQEDPLPLAFFVHDA 67  
QY 65 EIKKSLEDTLDLASVDTENVIDIVYQPOAVFKVRPVTRCTSSMPGHAEAVVSNFSPDGA 124  
Db 68 EIVSSLGKTLESQAVETEKVDIIYQPOAIFRVRAVTRCTSSLEGHSEAVISVAFSPTGK 127  
QY 125 HLASGSGDTTVRLWDLNTEPHFTCTGHKQWLCVSWAPDGKRLASGCKAGSIIWDPET 184  
Db 128 YLASGSGDTTVRFWDLSTETPHFTCKGHRHWVLSISWSPDGKKLASGCKNGQILLWDPST 187  
QY 185 GQOKRPLSGHKKHINCLAWEPYHRDPECRKLASASGDGDCRIWVKLGQCCLMNIAGHTN 244  
Db 188 GKQVGRTLAGHSKWITGLSWEPLHANPECRYVASSSKDGSVRIWDTTAGRCERILTGHTQ 247  
QY 245 AVTAVRWGGAGLIYTSKDRTVKMWRADGILCRTFSGHAHWVNNIALSTDYVLRGTGPFH 304  
Db 248 SVTCLRWGGDGLLYSASQDRTIKVWRAHDGVLCRTLOGHGHVWVWMLSTDYALRTGAFE 307  
QY 305 PVK-DRSKSHLSLSTEELQESALKRYQAVCPDEVELVSCDDNTLYLWRNNQN-KCVER 362  
Db 308 PAEASVNPQDLQGLQELKERALSRYNLVRGQGPRLVSGDDFTLFLWSPAEDKKPLTR 367  
QY 363 MTGHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDQGYMATFRGHVQAVYTVAWSADSR 422  
Db 368 MTGHQALINQVXFSPDSRIVASASFDSIKLWDGRTGKYLASLRGHVAAVYQIAWSADSR 427  
QY 423 LIVSGSKDSTLKVMSVQTKKLAQELPGHADEVFGVDWAPDGSRSVSGGKDKVIKWLW 478  
Db 428 LLVSGSSDSTLKVWMDVKAQKLAAMDLPGHADVEYAVDWSPDQQRVASGGGKDKCLRIW 483

RESULT 4  
AAB92844  
ID AAB92844 standard; Protein; 485 AA.  
XX  
AC AAB92844;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:11400.  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602





Db 308 PAEASVNPQDLQGSLOELKERALSRYNLVRGQPERLVSGSDDFTLFLWSPAEDKKPLTR 367  
QY 363 MTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDQYWATERGHVQAVYTVVAVSADSR 422  
Db 368 MTGHQALINQVLFSPDSRIVASAFDKSIKLWDGRTGKYLASLRGHVAAVYQIAWSADSR 427  
QY 423 LIVSGSKDSTLKVWSVQTKKLAQELPGHAEVFGVDWAPDGSRVASGKDKVIK1W 478  
Db 428 L1VSGSDSTLKVWDVKAQK1AMDLP1GHAEV1AVDWS1PDG1QRVAS1GKDK1CLR1W 483

RESULT 6  
ABB97306  
ID ABB97306 standard; Protein; 485 AA.  
XX  
AC ABB97306;  
XX  
DT 27-JUN-2002 (first entry)  
XX  
DE Novel human protein SEQ ID NO: 574.  
XX  
KW Human; antianaemic; vulnery; antinflammatory; immunomodulator;  
KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KW neuroprotective; antiparkinsonian; protein therapy; EST;  
KW expressed sequence tag.  
XX  
OS Homo sapiens.  
XX  
PN WO200222660-A2.  
XX  
PD 21-MAR-2002.  
XX  
PF 10-SEP-2001; 2001WO-US26015.  
XX  
PR 11-SEP-2000; 2000US-0659671.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX  
DR WPI; 2002-292408/33.  
DR N-PSDB; AEN32492.

An isolated polynucleotide for treating diseases associated with its encoded polypeptide such as cancer and multiple sclerosis -

Example 2; SEQ ID NO 574; 509pp; English.

The present invention provides the protein and coding sequences of 444 novel human proteins. These were isolated from expressed sequences tags (ESTs). They can be used to stimulate cell growth, to regulate haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth e.g. in burn treatment, to regulate the immune system e.g. to treat multiple sclerosis, to regulate activin or inhibin e.g. to treat infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke and cancer, to screen for drugs, to treat inflammatory conditions e.g. rheumatoid arthritis, and to treat nervous system disorders e.g. Parkinson's disease. The present sequence is a protein of the invention.

Sequence 485 AA;

Query Match 59.1%; Score 1520; DB 23; Length 485;  
Best Local Similarity 58.6%; Pred. No. 1.5e-134;  
Matches 279; Conservative 73; Mismatches 118; Indels 6; Gaps 3;

QY 9 EATPHTIQARLVYTGEA----GPPIDLPAGITTCQQLGLICNALLKNEATPPYLFVGED 64  
Db 8 EAVARDVQRLVQFQDEGGQLGSPFDVPVDITPDRQLQVNCALLAQEDPLPAFFVHDA 67  
QY 65 EIKKSLEDTLDLASVDTENVIDIVQPAVKVRPVRTCTSSMPGHAENVSLNFSPOGA 124  
Db 68 EIVSSLGKILESQAVETEKVLDIIYQPAIFRRAVTRCTSSLEGHSEAVISVAFSPGK 127

QY 125 HLASGSGDTTVRLWDLNLTETPHFTCTGHKQWLVCSWAPDGKRLASGCKAGSIIWDPET 184  
Db 128 YLASGSGDTTVRFWDLSTETPHFTCKGHRHWVLSISWSPDGKRLASGCKNGQILLWDPST 187  
QY 185 GOQGRPLSGHKKHINCLAWPEYHRDPECKRLASASGDGDCRIWDVKLGQCLMNIAGHTN 244  
Db 188 GKQVCRTLAGHSKWITGLSWEPLHANPECRYVASSSKDGSVR1WDTT1AGRCER1LT1GHTQ 247  
QY 245 AVTAVRWGGAGLIYTSKDRVYKMWRAADGILCRTEFSGHAHWVNNIALSTDYVLR1TGPFH 304  
Db 248 SVTCLRWGGDGLLYSASQDRTIKVWRAHDGVLCRTLQGHGHVWNTMALSTDYALRTGAFE 307  
QY 305 PVK-DRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDDNTLYLWRNNON-KCVER 362  
Db 308 PAEASVNPQDLQGSLOELKERALSRYNLVRGQPERLVSGSDDFTLFLWSPAEDKKPLTR 367  
QY 363 MTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDQYMATFRGHVQAVYTVVAVSADSR 422  
Db 368 MTGHQALINQVLFSPDSRIVASAFDKSIKLWDGRTGKYLASLRGHVAAVYQIAWSADSR 427  
QY 423 LIVSGSKDSTLKVWSVQTKKLAQELPGHAEVFGVDWAPDGSRVASGKDKVIK1W 478  
Db 428 L1VSGSDSTLKVWDVKAQK1AMDLP1GHAEV1AVDWS1PDG1QRVAS1GKDK1CLR1W 483

RESULT 7  
AAB68282  
ID AAB68282 standard; Protein; 484 AA.  
XX  
AC AAB68282;  
XX  
DT 09-JUL-2001 (first entry)  
XX  
DE Amino acid sequence of human Tzap gene cDNA clone Tzap7/B.  
XX

Tzap; T cell activation; immune response; transplant rejection; bone marrow transplantation; rheumatoid arthritis; lupus erythematosus; multiple sclerosis; encephalitis; vasculitis; diabetes mellitus; pancreatitis; gastritis; thyroiditis; malignant disorder; asthma; lepramatosus; gastritis; skin tumour; adrenal tumour; lung tumour; wound healing; growth disorder; inflammatory disease; infectious disease.

Homo sapiens.

WO200132614-A2.

10-MAY-2001.

30-OCT-2000; 2000WO-EP10670.

01-NOV-1999; 99US-0162675.

25-FEB-2000; 2000US-0185016.

(UTKU/) UTKU N.

Utku N;

WPI; 2001-316410/33.

N-PSDB; AAF85073.

Novel polynucleotide encoding Tzap protein involved in T cell activation, useful for diagnosing and treating diseases involving T cell activation, for treating organ transplantation rejection, rheumatoid arthritis -

Example; Page 49-51; 68pp; English.

The present sequence represents a Tzap protein. The Tzap gene is involved in T cell activation, and in the modulation of immune responses. Tzap polynucleotides and polypeptides are useful for diagnosing or treating acute and chronic diseases involving T cell activation and Th1 and Th2 immune response, for the treatment of acute

CC and chronic rejection of allo- and xeno organ transplants and bone  
CC marrow transplantation, for the treatment of rheumatoid arthritis, lupus  
CC erythematosis, multiple sclerosis, encephalitis, vasculitis, diabetes  
CC mellitus, pancreatitis,gastritis, thyroiditis, for the treatment of  
CC malignant disorders of T, B or NK cells, for the treatment of asthma,  
CC lepramatosiis, Helicobacter pylori associated gastritis or for the  
CC treatment of skin tumours, adrenal tumours or lung tumours, wound  
CC healing, growth disorders inflammatory and/or infectious diseases.  
XX  
SQ Sequence 484 AA;

Query Match 56.4%; Score 1452; DB 22; Length 484;  
Best Local Similarity 57.6%; Pred. No. 4e-128;  
Matches 274; Conservative 66; Mismatches 130; Indels 6; Gaps 3;

QY 9 EATPHTIQARLVYTGEEA---GPPIDLPGITTTQQLGICNALLKNEATPYLFFVGED 64  
Db 7 EAVARDVQRLLVQFQDEGGQLLSPFDVPVDITPDRLLQVLNALLAQEDPLSLAFFVHDA 66  
QY 65 EIKKSLEDTLDLASVDTENVIDIVYQPAVKVRPVTRCTSSMPGHAEAVVSLNESPdGA 124  
Db 67 EIVSSLGKLTLESQAVETEKVLDIILPTQAVFKVRAVTRCTSSLEGHTEAVISVAFSPtGK 126  
QY 125 HLASGSGDTTVRLWDLNLTETPHFTCTGHKQWVLCVSWAPDGKRLASGCKAGSIiIWDPET 184  
Db 127 YLASGSGDTTVRFWDLSTETPHFTSKGHTHWVLSIAWSPDGKKLASGCKNSQIFIWDPST 186  
QY 185 GQKGRPLSGHKKHINCLAWEPYHRDPECRKILASASGDGDCRIWDVKLGQCLMNIAGHTN 244  
Db 187 GKQIGKPLTGHSKWITWLCWEPLHNPESRYLASASSGRYDRIWDTTAGRCERILTGHTQ 246  
QY 245 AVTAVRWGGAGLIYTSSKDRTVKMWRAADGILCRTFSGHAHWNNIALSTDYVLTGPpFH 304  
Db 247 SVTCLRWGGDLLYSASQDRTIKVRAHDGVLCRTLQHGCHWVNTMALSTDYALRTGAFE 306  
QY 305 PVK-DRSKSHLSLSTEELOESALKRYQAVCPDEVESLVSCDDNTLLILWRNNQN-KCVER 362  
Db 307 PAEASVNPQDLQGLQELKERALSRYNLVRGQGPRLVSGSDDFTFLWSPAEDKKPLTR 366  
QY 363 MTGHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMATFRGHVQAVYTVAMSADSR 422  
Db 367 MTGHQALINQVLFSPDSRIVASAFDKSIKLWDGRTGKYLASLRGHVAAVYQIAWSADSR 426  
QY 423 LIVSGSKDSTLKVWSVQTKKLAQLPGHADEVGVDWAPDGSRVASGCKDKVIKLW 478  
Db 427 LLVSGSSDSTLKVVDVYKAQKLA MDLPGHADEVYAVDWSPDGQRVASGCKDKCLRiW 482

RESULT 8  
AAG231141  
ID AAG23141 standard; Protein; 471 AA.  
XX  
AC AAG23141;

DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 26337.  
DE Arabidopsis thaliana.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.

PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 51.9%; Score 1335.5; DB 21; Length 471;  
Best Local Similarity 52.5%; Pred. No. 3.9e-117;  
Matches 256; Conservative 73; Mismatches 130; Indels 29; Gaps 6;

QY 1 MQETD-----TEQEATPHTIQARLVYTGEEAGPPIDLPAGITTOQLGLICNALLKN 51  
Db 1 MMDTDEGKIVMCLLTDPGT-HLGSA--MYIPQKAGP-----LQTLQLVNRFELDN 47  
QY 52 EEATPYLFFVGEDEIKKSLEDTLDLASVDTENVIDIVYQPAVKVPRVTRCTSSMPGHA 111  
Db 48 EEMPLYSFYVSDELLVPVGTYLEKNKVSVEKVLTVYQQQAVFIRIRPVNRCSTIAGHA 107  
QY 112 EAVVSLNFPDGAHLASGDDTTVRLWDLNLTETPHFTCTGHKQWVLCVSWAPDGKRLASG 171  
Db 108 EAVLCVSFPDGGKQLASGDDTTVRLWDLNLTETPLCTCKGHKNWVLTVAWSPDGKHLVSG 167  
QY 172 CKAGSIITWDPETGOQKGRPLSGHKKHINCLAWEPYHRDPECKKLASASGDCRWDVK 231  
Db 168 SKSGEICCNPKKGELEGSPLTGHKKWITGISWEPVHLSSPCRFVTSSKGDARIDIT 227  
QY 232 LGQCLMNIAGHTNAVTVRWGGAGLIYTSSKDRYVKMWRAADGILCRTFSGHAHWNNIA 291  
Db 228 LKKSIIICLSGHTLAVTCVKWGGDIYTGSDCTIKMWETTQGLIRELKGHGHWINSLA 287  
QY 292 LSTDYVLRTPGFHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDNTLYL 351  
Db 288 LSTEYVLRGTAF-----DHTGRQYPPNEEKQKALERYNKTKGDSPERLVSGSDDTMFL 341  
QY 352 WRNNQNK-CVERMTGHNQVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMATFRGHVQ 410  
Db 342 WEPVSVKOPKKRLTGHQOLNVHNYFSPDGKWIASAFDKSVRLWNGITGQFVTVFRGHVG 401  
QY 411 AVYTVAWSADSRLLIVSCSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWADPGSRVASGG 470  
Db 402 PVYQVWSADSRLLLSGSKDSTLKIWEIRTKKLQDLPGHADVFVADWSPDGEKVVSGG 461  
QY 471 KDKVIKWL 478  
Db 462 KDRVLKLW 469

RESULT 9  
AAG46521  
ID AAG46521 standard; Protein; 471 AA.  
XX  
AC AAG46521;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 58536.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.



PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 51.9%; Score 1335.5; DB 21; Length 471;		
Best Local Similarity 52.5%; Pred. No. 3.9e-117;		
Matches 256; Conservative 73; Mismatches 130; Indels 29; Gaps 6;		
QY	1 MQETD-----TEQETPHTTQARLVYTGEEAGPPIDLPAGITTQQLGLICNALLKN 51	
Db	1 MMDTDEGKTVMCLLTDPEGT-HLGSA--MYIPQAGP-----LQLTQLVNRFLDN 47	
QY	52 EEATPYLFFVGEDEIKKSLEDTLDLASVDTENVIDIVYQPOAVFKVRPVTRCTSSMPGHA 111	
Db	48 EEMPLPYSFYVSDSELLVPVGTYLEKNKVSVEKVLTIYQQQAVFRIRPVNRCSTIAGHA 107	
QY	112 EAVVSLNFPSPDGAHLASGSGDTTVRLNDLNTETPHFTCTGCHKQWVLCVSWAPDGKRLASG 171	
Db	108 EAVLCVSFSPDGKQLASGSGDTTVRLNDLTYETPLFTCKGHNWVLTVAWSPDGKHLVSG 167	

QY	172 CKAGSIIIWDPETGQKGRPLSGHKKHINCLAWPEPYHRDPECKKLASASGDCRWDVK 231	
Db	168 SKSGEICWNPKKGELEGSPLTGHRKKWITGISWEPVHLSSPCRRFVTSKDGDAWIWIT 227	
QY	232 LGQCLMNIAGHTNAVAVRWGGAGLIYTSSKDRVTKMWRADGILCRTLTFSGHAHWNNIA 291	
Db	228 LKKSIIICLSGHTLAVTCVKGWGGDIYTGSDCTIKMWETTQKLIRELKKGHWINSLA 287	
QY	292 LSTDYVLRTPFHPVKDRSKSHLSLSTEELOESALKRYQAVCPDEVESLVSCSDDNTLYL 351	
Db	288 LSTEYVLRTGAF-----DHTGRQYPPNEEKQKALERYNKTKGDSPERLVSGSDDFTMFL 341	
QY	352 WRNNQNK-CVERMTGHONVNDVKYSPDVVKLIASASFDKSVRLWRASOGQYMATFRGHVQ 410	
Db	342 WEPSVSKQPKRLTGHQQLVNHVYFSPDGKWIASASFDKSVRLWNGITGQFVTVFRGHVG 401	
QY	411 AVYTVANSADSRLLIVSGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWADGSRVASGG 470	
Db	402 PVYQVWSADSRLLLSGSKDSTLKIWEIRTKKLKQDLPGHADEVFAVDWSPDGEKVVSGG 461	
QY	471 KDKVIKLW 478	
Db	462 KDRVCLKW 469	
RESULT 10		
ID	AAG23140	
XX	AAG23140 standard; Protein; 473 AA.	
AC	AAG23140;	
XX	17-OCT-2000 (first entry)	
DT		
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 26336.	
DE		
XX	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX	Arabidopsis thaliana.	
OS		
XX	EP1033405-A2.	
PN		
XX	06-SEP-2000.	
PD		
XX	25-FEB-2000; 2000EP-0301439.	
PF		
XX	99US-0121825.	
PR	99US-0123180.	
PR	99US-0123548.	
PR	99US-0125788.	
PR	99US-0126264.	
PR	99US-0126785.	
PR	99US-0127462.	
PR	99US-0128234.	
PR	99US-0128714.	
PR	99US-0129845.	
PR	99US-0130077.	
PR	99US-0130449.	
PR	99US-0130510.	
PR	99US-0130891.	
PR	99US-0131449.	
PR	99US-0132048.	
PR	99US-0132407.	
PR	99US-0132484.	
PR	99US-0132485.	
PR	99US-0132486.	
PR	99US-0132487.	
PR	99US-0132863.	
PR	99US-0134256.	
PR	99US-0134218.	
PR	99US-0134219.	
PR	99US-0134221.	

PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.

```
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 51.9%; Score 1335.5; DB 21; Length 473;
Best Local Similarity 52.5%; Pred. No. 3.9e-117;
Matches 256; Conservative 73; Mismatches 130; Indels 29; Gaps 6;

QY 1 MQETD-----TEQEARPTHTIOARLVYTGEAGPPIDLPAGITTQOLGLICNALLKN 51
Db 3 MMDTDEGKTVMCLLTDPEGT-HLGSA--MYIPQKAGP-----LQLTQLVNRFLDN 49

QY 52 EEATPYLFFVGEDEIKKSLEDHDLASVDTENVIDIVYQPAVFKVRPVRTCTSSMPGHA 111
Db 50 EEMPLYSFYVSDDELLVPVGYLEKNKVSVEKVLTIYVQQQAVFRIRPVNRCSTIAGHA 109

QY 112 EAVVSLNFSPDGAHLASGSDTIVRLWDLNTETPHFTCTGKHQWVLCVSWAPDGKRLASG 171
Db 110 EAVLCVSFSPDGKQLASGSDTIVRLWDLTYETPLFTCKGKHNWVLTVAWSPDGKHLVSG 169

QY 172 CKAGSIIWDPETGQOKGRPLSGHKKHINCLAWEPYHRDPECKRLASASGDGDCRIWDVK 231
Db 170 SKSGEICCWNPKKGELEGSPLTGHKKWITGTSWEPVHLSSPCRRFVTSKDKGDARIWDIT 229

QY 232 LGQCLMNIAGHTNAVTAVRWGGAGLIYTSKDRVTVMKMRADGILCFTFSGHAWVNNIA 291
Db 230 LKKSIIICLSGHTLAVTCVKWGGDGIIYTGSDCTIKMWETTQKKLIRELKGHGWINSLA 289

QY 292 LSTDYVLRTPGFHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDNTLYL 351
Db 290 LSTEYVLRRTGAF-----DHTGRQYPPNEKQKALERYNKTKGDSPERLVSGSDFTMFL 343

QY 352 WRNNQNK-CVERMTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMAFRGHVQ 410
Db 344 WEPVSQKQPKKRLTGHQQLVNHVYFSPDGKWIASAFDKSVRLWNGITGQVTVFRGHVG 403

QY 411 AVYTVAMSADSRLLIVSGSKDSTLKVMSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGG 470
Db 404 PVYQVWSADSRLLSGSKDSTLKIWEIRTKKQLKQDLPGHADEVFAVDWSPDGKRVVSGG 463

QY 471 KDKVIKILW 478
Db 464 KDRVCLKLW 471

RESULT 11
AAG46520
ID AAG46520 standard; Protein; 473 AA.
XX AC AAG46520;
XX AC AAG46520;
DT 18-Oct-2000 (first entry)
XX DE
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58535.
XX DE
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX OS
PN EP1033405-A2.
XX PD
XX PD 06-SEP-2000.
XX PF
XX PF 25-FEB-2000; 2000EP-0301439.
XX PF
PR 25-FEB-1999; 99US-0121825.
```

```
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
```





RESULT 12  
AAG23139  
ID AAG23139 standard; Protein; 490 AA.  
XX  
AC  
XX  
AC AAG23139;  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 26335.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.



PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 51.9%; Score 1335.5; DB 21; Length 490;  
Best Local Similarity 52.5%; Pred. No. 4.1e-117;  
Matches 256; Conservative 73; Mismatches 130; Indels 29; Gaps 6;

QY	1	MQETD-----TEQEATPHITQARLVYTGEAGPPIDLPAGITTOQLGLICNALUKN	51
Db	20	MMDTDEGKTMCLLTDPEGT-HLGSA--MYIPQAGP-----LQITQLVNRFDLN	66
QY	52	EEATPYLFFVGEDETKKSLEDITLDLASVDTENVIDIVYQPAQVFKVRPVTRCTSSMPGHA	111
Db	67	EEMLPYSFYVSDEELLVPVGTYLEKNKVSVEKVLITIVYQQQAVFIRIPVNRCSQTIAGHA	126
QY	112	EAVSLNFSPDGAHLASGSGDITVRLWDLNTEPHFTCTGHKQWVLCVSWAPDGKRLASG	171
Db	127	EAVLCVSFSPDGKQLASGSGDITVRLWDLTYETPTLFTCKGHKNWVLTVAWSPDGKHLVSG	186
QY	172	CKAGSIIDWPETGOOKGRPLSGHKHKNCLAWEPYHRDPECRKKLASAGDGDCCRINDVK	231
Db	187	SKSGEICCWNPKGEBLGSPLTGHKKWITGISWEPVHLSSPCRRFVTSKDGDDARIWDIT	246
QY	232	LGQCLMNIAGHTNAVTVRWGGAGLIYTSKDBTVKMWRADGILCRTTSGHAHVNNITA	291
Db	247	LKKSIIICLSGHTLATVCVKWGGDIIYTGSDQCTIKMWETTQGLIRELKGGHGWINSIA	306
QY	292	LSTDYVLRGTGPFHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVCSDDNLTLYL	351
Db	307	LSTEYVLRGTAF-----DHTGRQYPPNEEKQKALERYNKTKGDSPERLYSGSDDETFML	360
QY	352	WRNNQNK-CVERMTGHQNVNDVYKSPDVKLIASASFDKSVRLWRASDGOYMATFRGHVQ	410
Db	361	WEPVSQPKPKRLTGHQQLVNVHYFSPDGKWJTASASFDKSVRLWNNGITGQFVTVFRGHVG	420
QY	411	AVYTVAWSADSRLLIVSGSKDSTLKWVSVQTKLAQELPGHADEVFGVDWAPDGSRRVASGG	470
Db	421	PVYQVSWASADSRLLISGSKDSTLKIWEIRTKKLKQDLPGHADDEVFAVDWSPDGKVVSG	480
QY	471	KDKVIKWL 478	
Db	481	KDRVKLWL 488	

RESULT 14  
AAR85881  
ID AAR85881 standard; Protein; 514 AA.

AC AAR85881;

DT 13-SEP-1996 (first entry)

DE WD-40 domain-contg. YCW2 p

KW WD40 repeat region; beta-transduc

receptors of activated protein kinase; enzyme activity; isozyme; human.

OS Synthetic.

PN W09521252-A2.

PD 10-AUG-1995.

PF 31-JAN-1995; 95WO-US01210.

01-FEB-1994; 94US-0190802.  
(STRD ) UNIV LELAND STANFORD JUNIOR.  
Mochly-Rosen D, Ron D;  
WPI; 1995-283772/37.  
New WD-40 (beta-transducin)-derived polypeptide(s) - which alter the activity of a protein, eg. protein kinase C, which interacts with a protein contg. a WD-40 region.  
Example 5; Page 168-170; 351pp; English.  
Proteins AAR5851-92 are protein which contain at least one WD-40 (also called beta-transducing homologous) amino acid repeat motifs. The WD-40 regions are involved in protein-protein interactions between proteins involved in intracellular signalling. An example of such an interaction is between protein kinase C and receptors of activated protein kinase (RACK), esp. RACK-1 (AAR85850). Proteins AAR5851-82 were isolated based on homology with beta-transducin, whereas proteins AAR5882-92 were isolated based on homology with the WD-40 consensus sequence (AAR5893). The proteins were used to construct the peptides AAR84928-R85063 and AAR85786-R85842. The peptides can be used to identify target proteins contg. WD-40 motifs, as modulators of enzyme esp. isozyme, activity of proteins involved in protein-protein interaction and to screen for drugs that will affect protein-protein interaction involving WD-40 domains.

QY	14	TIQARLVYTGEAEAGPIDLPAGITTOQLGLICNALL-KNEEATPYLPFV-----GEDEI	66
Db	33	SIKFOALDTGDNVGGALRVPGAISEKQLEELLNQLNGTSDDDPVPTTSCITIOGKKASDPV	92
QY	67	KKSLEDITLDAS-----VDTENVIDIVYQPAQVFKVRPVTRCTSSMPGHAERAVV	115
Db	93	K-----TIDITDNLSSLLIKPGYNSTEDOITLLYTTPRAVFKVKPVTRSSSAIAGHGSTIL	147
QY	116	SLNFSP-DGAHLASGSGDPTVRLWDLNTEPHFTCTGHKQWVLCVSWAPDGKRLASGCKA	174
Db	148	CSAFAPHTSSRMVGTAGDNTARIWDCDTQTPMHTLKGHYNWVLCVSWSPDGEVIATGSM	207
QY	175	GSIIWDPETGQKGRPLSGHKKHINCLAWEPYH--RDPECRKKLASAGDGDCCRINDVKL	232
Db	208	NTIRLWDPKSGQCLGDALRGHSKWIITSLWEPIHLVAPGSKPRLASSKSDGTIKIWDTVS	267
QY	233	GQCLMNTAGHTNAVTAVRWGGAGLIYTSKDRTVKMWRA-ADGILCRTEFSGHAHWVNIA	291
Db	268	RVCQYTMSGHTNSVSCVKWGGQGLLYSGSHDRTVRVWDINSQGRGINILKSHAHWVNHL	327
QY	292	LSTDYVLRGTPF-HPVDRSKSHLSLSTEELQESALKRYQAVCP---DEVESLVSCSDDN	347
Db	328	LSTDYALRIGAFDHTGKRPS-----TPEEAQKKALENYEKICKKNSEEMMVTASDDY	381
QY	348	TLYLWRN-NQNKCVERTMTHQNVNDVKYSPDVVKLIASAFDKSVRLWRASDGOYMATFR	406
Db	382	TMFLWNPLKSTKPIARMTGCHOKLVNHVAFSPDGRYIVTSASFDSNIKLDGRDGKFIFR	441
QY	407	GHVQAVITYAWSADSRLLIYSGSKDSTLKVWSVOTKLQAQELPGHADEVFGVDWAPDGSRV	466
Db	442	GHIASVYQVAWSSDCRLLYVSCSKDTTLKVWDVTRKLSVDLPGIKTKLY-VDWSVDGKRV	500
QY	467	ASGGKDKVIKLWAY	480
Db	501	CSGGKDKMVRLLWTH	514

## RESULT 15







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:58:20 ; Search time 35 Seconds  
(without alignments)  
403.514 Million cell updates/sec

Title: US-09-830-980-1  
Perfect score: 2574  
Sequence: 1 MQETDEQATPHTIQARLV.....PDGSRVASGGDKVKIKLWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1058.5	41.1	514	1	US-08-190-802A-66
2	1058.5	41.1	514	4	US-08-477-346-66
3	1058.5	41.1	514	4	US-08-473-089-66
4	1058.5	41.1	514	4	US-08-487-072A-66
5	383.5	14.9	251	4	US-09-291-170A-13
6	383.5	14.9	251	4	US-09-724-884-13
7	354	13.8	409	2	US-08-283-917-3
8	354	13.8	409	2	US-08-961-716-3
9	354	13.8	410	2	US-08-283-917-9
10	354	13.8	410	2	US-08-961-716-9
11	340.5	13.2	409	1	US-08-190-802A-51
12	340.5	13.2	409	4	US-08-477-346-51
13	340.5	13.2	409	4	US-08-473-089-51
14	340.5	13.2	409	4	US-08-487-072A-51
15	323.5	12.6	343	4	US-09-063-743-5
16	323.5	12.6	343	4	US-09-590-540-5
17	312	12.1	1194	4	US-09-092-508-2
18	312	12.1	1194	4	US-09-435-115-2
19	312	12.1	1194	4	US-09-069-023-26
20	312	12.1	1194	4	US-09-098-310-2
21	312	12.1	1205	4	US-09-092-508-16
22	312	12.1	1205	4	US-09-435-115-16
23	310.5	12.1	640	4	US-09-177-165A-30
24	308	12.0	375	4	US-09-063-743-1
25	308	12.0	375	4	US-09-590-540-1
26	306	11.9	517	1	US-08-190-802A-30
27	306	11.9	517	4	US-08-477-346-30

28	306	11.9	517	4	US-08-473-089-30	Sequence 30, Appl
29	306	11.9	517	4	US-08-487-072A-30	Sequence 30, Appl
30	305	11.8	514	4	US-09-108-857-2	Sequence 2, Appl
31	302.5	11.8	2627	2	US-08-751-189-3	Sequence 3, Appl
32	302.5	11.8	2627	2	US-09-060-836-3	Sequence 3, Appl
33	302.5	11.8	2627	4	US-09-184-445-3	Sequence 3, Appl
34	297	11.5	587	3	US-08-899-578-2	Sequence 2, Appl
35	294.5	11.4	713	1	US-08-190-802A-63	Sequence 63, Appl
36	294.5	11.4	713	4	US-08-477-346-63	Sequence 63, Appl
37	294.5	11.4	713	4	US-08-473-089-63	Sequence 63, Appl
38	294.5	11.4	713	4	US-08-487-072A-63	Sequence 63, Appl
39	294.5	11.4	713	4	US-09-108-857-3	Sequence 3, Appl
40	294	11.4	253	4	US-09-291-170A-10	Sequence 10, Appl
41	294	11.4	253	4	US-09-724-884-10	Sequence 10, Appl
42	294	11.4	690	4	US-09-291-170A-2	Sequence 2, Appl
43	294	11.4	690	4	US-09-724-884-2	Sequence 2, Appl
44	292.5	11.4	404	4	US-09-184-001-2	Sequence 2, Appl
45	291	11.3	704	1	US-08-188-582-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1  
US-08-190-802A-66  
; Sequence 66, Application US/08190802A  
; Patent No. 5519003  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: P.O. Box 60850  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-0850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/190,802A  
; FILING DATE: 01-FEB-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fabian, Gary R.  
; REGISTRATION NUMBER: 33,875  
; REFERENCE/DOCKET NUMBER: 8600-0139  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 66:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 514 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: YCW2 PROTEIN, Fig. 49  
; US-08-190-802A-66

Query Match 41.1%; Score 1058.5; DB 1; Length 514;  
Best Local Similarity 42.5%; Pred. No. 8.5e-98;  
Matches 210; Conservative 97; Mismatches 148; Indels 39; Gaps 12;  
QY 14 TIQARLVYTGEEAGPPIDLPAGITTTQQLGLICNALL-KNEEATPYLFFV-----GEDEI 66

Db 33 SIKFQALDTGDNVGGALRVPGAISEKLELLNQLNGTSDDPVPTFSCITQGGKASDPV 92  
QY 67 KKSLEDTLDLAS-----VDPENVIDIVYQPAVEKVPVTRCTSSMPGHAEEAVV 115  
Db 93 K-----TIDITDNLSSLIKPGYNSTEDQITLLYTPRAVEKVPVTRSSAIAHGSGTIL 147  
QY 116 SLNFSP-DGAHLASGSGDTTVRLWDLNLTETPHFTCTGHKQWVLCVSWAPDGKRLASGCKA 174  
Db 148 CSAFAPHTSSRMVTGAGDNTARIWDCDTQTPMHTLKGHYNWVLCVSWSPDGEVIATGSMD 207  
QY 175 GSIIWDPETGQKGRPLSGHKKHINCLAWEPYH--RDPECKRLASASGDGDCRIWDVKL 232  
Db 208 NTIRLWDPKSGQCLGDALRGHSKWITSLWEPHILVKPGSKPRLASSSKDKGTIKIWDTVS 267  
QY 233 GQCLMNIAGHTNAVAVRWGGAGLIYTSKSDRTVKMWRA-ADGILCRTFSGHAWVNNIA 291  
Db 268 RVCQYTMGHTNSVSCVKWGGQGLLYSGSHDRTVRVWDINSQGRGINILKSHAHVNNHLS 327  
QY 292 LSTDYVLRTPGF-HPVKDRSKSHLSSTEELESALKRYQAVCP--DEVESLVSCSDDN 347  
Db 328 LSTDYALRIGAFDHTGKKPS-----TPEEAQKALENYEKICKKNGNSEMMVTASDDY 381  
QY 348 TLYLWRN-NQKCVERTGHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMATER 406  
Db 382 TMFLWNLKSTKPIARMTGHQKLVNHVAFSPDGRYIVSASFONSIKLWDRGDKGFISTFR 441  
QY 407 GHVQAVYTVAWSADSRLLIVSGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRV 466  
Db 442 GHIASVYQVAWSSDCRLVSCSKDTTLKVWDVTRKLSVDLPGIKTKLY-VDWSVDGKRV 500  
QY 467 ASGGKDKVIKLWAY 480  
Db 501 CSQKDKMVRWLWTH 514

RESULT 2

US-08-477-346-66  
; Sequence 66, Application US/08477346  
; Patent No. 6262023  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477,346  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/487,072  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2550-0025.20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; INFORMATION FOR SEQ ID NO: 66:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 514 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: YCW2 PROTEIN, Fig. 49  
US-08-477-346-66  
  
Query Match 41.1%; Score 1058.5; DB 4; Length 514;  
Best Local Similarity 42.5%; Pred. No. 8.5e-98;  
Matches 210; Conservative 97; Mismatches 148; Indels 39; Gaps 12;  
  
QY 14 TIQARLVYTGEEAGPPIDLPAGITTOQLGLICNALL-KNEEATPYLFFV-----GEDEI 66  
Db 33 SIKFQALDTGDNVGGALRVPGAISEKLELLNQLNGTSDDPVPTFSCITQGGKASDPV 92  
QY 67 KKSLEDTLDLAS-----VDTENVIDIVYQPAVEKVPVTRCTSSMPGHAEEAVV 115  
Db 93 K-----TIDITDNLSSLIKPGYNSTEDQITLLYTPRAVEKVPVTRSSAIAHGSGTIL 147  
QY 116 SLNFSP-DGAHLASGSGDTTVRLWDLNLTETPHFTCTGHKQWVLCVSWAPDGKRLASGCKA 174  
Db 148 CSAFAPHTSSRMVTGAGDNTARIWDCDTQTPMHTLKGHYNWVLCVSWSPDGEVIATGSMD 207  
QY 175 GSIIWDPETGQKGRPLSGHKKHINCLAWEPYH--RDPECKRLASASGDGDCRIWDVKL 232  
Db 208 NTIRLWDPKSGQCLGDALRGHSKWITSLWEPHILVKPGSKPRLASSSKDKGTIKIWDTVS 267  
QY 233 GQCLMNIAGHTNAVAVRWGGAGLIYTSKSDRTVKMWRA-ADGILCRTFSGHAWVNNIA 291  
Db 268 RVCQYTMGHTNSVSCVKWGGQGLLYSGSHDRTVRVWDINSQGRGINILKSHAHVNNHLS 327  
QY 292 LSTDYVLRTPGF-HPVKDRSKSHLSSTEELESALKRYQAVCP--DEVESLVSCSDDN 347  
Db 328 LSTDYALRIGAFDHTGKKPS-----TPEEAQKALENYEKICKKNGNSEMMVTASDDY 381  
QY 348 TLYLWRN-NQKCVERTGHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMATER 406  
Db 382 TMFLWNLKSTKPIARMTGHQKLVNHVAFSPDGRYIVSASFONSIKLWDRGDKGFISTFR 441  
QY 407 GHVQAVYTVAWSADSRLLIVSGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRV 466  
Db 442 GHIASVYQVAWSSDCRLVSCSKDTTLKVWDVTRKLSVDLPGIKTKLY-VDWSVDGKRV 500  
QY 467 ASGGKDKVIKLWAY 480  
Db 501 CSQKDKMVRWLWTH 514

RESULT 3

US-08-473-089-66  
; Sequence 66, Application US/08473089  
; Patent No. 6342368  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS





```
QY 467 ASGGKDKVIKLMAY 480
Db 501 CSGGKDKMVLWTH 514

RESULT 5
US-09-291-170A-13
; Sequence 13, Application US/09291170A
; Patent No. 6410687
; GENERAL INFORMATION:
; APPLICANT: Vale, Ronald D.
; APPLICANT: Hartman, James J.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Assays for the Detection of Microtubule
; TITLE OF INVENTION: Depolymerization Inhibitors
; FILE REFERENCE: 18557B-000510US
; CURRENT APPLICATION NUMBER: US/09/291,170A
; CURRENT FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: US 60/081,734
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Thermomonospora curvata
; FEATURE:
; OTHER INFORMATION: putative serine/threonine kinase pKwA WD40 repeat
; OTHER INFORMATION: region
US-09-291-170A-13

Query Match 14.9%; Score 383.5; DB 4; Length 251;
Best Local Similarity 26.1%; Pred. No. 2.2e-30;
Matches 87; Conservative 50; Mismatches 107; Indels 89; Gaps 3;

QY 106 SMPGHAEAVVSLNFSPDGAHLASGSDTTVRLWDLNLTETPHTCTGKHQWVLCVSWAPDG 165
Db 8 TLEGHTDWVRVAVAFSPDGALLASGSDDATVRLWDVAAAEEERAVFEGTHYVLDIAFSPDG 67
QY 166 KRLASGCKAGSIIWDPETGQOKRPLSGHKKHINCLAWEPYHRDPECKRLASASGDGDC 225
Db 68 SMVASGSRDGTARLWNVATGTEHA-VLKGHTDYVYAVAF-----SPDGMVASGSRDGTI 121
QY 226 RIWDVKLGQCLMNIAGHTNAVTAVRWGGAGLIYTSSKDRTVKMWRAADGILCRTFSGHAH 285
Db 122 RLWDVATGKERDVLQAPAEENVVSLAFSPDGSMLVHGSDSTVHLWDVASGEALHTFEGHTD 181
QY 286 WYNNIALSTDYVLRTPGPHFPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSD 345
Db 182 WVRAVA----- 187
QY 346 DNTLYLWRNNQKCVERTMTHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMATF 405
Db 188 -----FSPDGALLASGSDDTIRLWDVAAQEEHTTL 218
QY 406 RGHVQAVYTVAWSADSRILVSGSKDSTLKVWSV 438
Db 219 EGHTPEVHSVAFHPEGTTLASASEDGTIRIWIPI 251

RESULT 6
US-09-724-884-13
; Sequence 13, Application US/09724884
; Patent No. 6429304
; GENERAL INFORMATION:
; APPLICANT: Vale, Ronald D.
; APPLICANT: Hartman, James J.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Assays for the Detection of Microtubule
; TITLE OF INVENTION: Depolymerization Inhibitors
; FILE REFERENCE: 18557B-000510US
; CURRENT APPLICATION NUMBER: US/09/724,884
; CURRENT FILING DATE: 2000-11-28
```

```
; PRIOR APPLICATION NUMBER: 09/291,170
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Thermomonospora curvata
; FEATURE:
; OTHER INFORMATION: putative serine/threonine kinase pKwA WD40 repeat
; OTHER INFORMATION: region
US-09-724-884-13

Query Match 14.9%; Score 383.5; DB 4; Length 251;
Best Local Similarity 26.1%; Pred. No. 2.2e-30;
Matches 87; Conservative 50; Mismatches 107; Indels 89; Gaps 3;

QY 106 SMPGHAEAVVSLNFSPDGAHLASGSDTTVRLWDLNLTETPHTCTGKHQWVLCVSWAPDG 165
Db 8 TLEGHTDWVRVAVAFSPDGALLASGSDDATVRLWDVAAAEEERAVFEGTHYVLDIAFSPDG 67
QY 166 KRLASGCKAGSIIWDPETGQOKRPLSGHKKHINCLAWEPYHRDPECKRLASASGDGDC 225
Db 68 SMVASGSRDGTARLWNVATGTEHA-VLKGHTDYVYAVAF-----SPDGMVASGSRDGTI 121
QY 226 RIWDVKLGQCLMNIAGHTNAVTAVRWGGAGLIYTSSKDRTVKMWRAADGILCRTFSGHAH 285
Db 122 RLWDVATGKERDVLQAPAEENVVSLAFSPDGSMLVHGSDSTVHLWDVASGEALHTFEGHTD 181
QY 286 WYNNIALSTDYVLRTPGPHFPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSD 345
Db 182 WVRAVA----- 187
QY 346 DNTLYLWRNNQKCVERTMTHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMATF 405
Db 188 -----FSPDGALLASGSDDTIRLWDVAAQEEHTTL 218
QY 406 RGHVQAVYTVAWSADSRILVSGSKDSTLKVWSV 438
Db 219 EGHTPEVHSVAFHPEGTTLASASEDGTIRIWIPI 251

RESULT 7
US-08-283-917-3
; Sequence 3, Application US/08283917
; Patent No. 5849557
; GENERAL INFORMATION:
; APPLICANT: ADACHI, HIDEKI
; APPLICANT: TSUJIMOTO, MASAFUMI
; APPLICANT: INOUE, KEIZO
; APPLICANT: ARAI, HIROYUKI
; TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME
; TITLE OF INVENTION: AND GENE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &
; ADDRESSEE: NEUSTADT,P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/283,917
; FILING DATE: 03-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 209943/1993
```

```

; FILING DATE: 03-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5849557man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2292-030-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
;
US-08-283-917-3

Query Match 13.8%; Score 354; DB 2; Length 409;
Best Local Similarity 24.1%; Pred. No. 4.5e-27;
Matches 85; Conservative 65; Mismatches 135; Indels 68; Gaps 5;

QY 147 FTCTGHKQWVLCVSWAPDGKRLASGCKAGSIIIWDPETGQKGRPLSGHKKHINCLAWEP 206
Db :||:| | | | | :||:| ||| :| | | | :||:| | | :||:| | |
101 YALSGHRSPVTRVIFHPVFSVMVSASEDATIKVWDYETGDFE-RTLKGHTDSVQDISF-- 157

QY 207 YHRDPECRKLASAGDGCRIWVVKLGQCLMNIAGHTNAVTAVR-WGGAGLIYTSSKDR 265
Db | :||| | | | :||:| | :||:| | | :||:| | | :||:| | |
158 ---DHSGKLLASCSADMTIKLWDFQGFECIRTMHGHDNVSSVAIMPNGDHIYSASRDKT 214

QY 266 VKMWRADGILCRTFSGHAHVNNIALSTDYVLRTPGPHVPVKDRSKSHLSLSTEELQESA 325
Db :||| | :||:| | | :| | | :| | | :| | | :| | |
215 IKMWEVQTGYCVKFTFGHREWVRMVRPNQDGTL----- 247

QY 326 LKRYQAVCPDEVESILVSCDDNTLYLWRNNQKCVERTGHQNVNDVKYSPDYK----- 380
Db :||:| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
248 -----IASCSNDQTVRVVVVATKECKAELREHEHVVECISSWAPESYSSIS 293

QY 381 -----LIASAFDKSVRLWRASDGOYMATFRGHVQAVYTVAWSADSRLLIV 425
Db :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
294 EATGSETKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNVRGVLFHSGGKFIL 353

QY 426 SGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKWL 478
Db | :| | | | :| :| | | :| :| :| :| :| :| :| :| :| :| :| :| :|
354 SCADDKTLRVWDYKNKRCMKTLNAHEHFTSLDFHKHTAPYVVTGSDVDTVKVW 406
```

```

RESULT 8
US-08-961-716-3
; Sequence 3, Application US/08961716
; Patent No. 5880272
; GENERAL INFORMATION:
; APPLICANT: ADACHI, HIDEKI
; APPLICANT: TSUJIMOTO, MASAFUMI
; APPLICANT: INOUE, KEIZO
; APPLICANT: ARAI, HIROYUKI
; TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME
; TITLE OF INVENTION: AND GENE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &
; ADDRESSEE: NEUSTADT,P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,716
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/283,917
; FILING DATE: 03-AUG-1994
; APPLICATION NUMBER: JP 209943/1993
; FILING DATE: 03-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5880272man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2292-030-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
;
US-08-961-716-3

Query Match 13.8%; Score 354; DB 2; Length 409;
Best Local Similarity 24.1%; Pred. No. 4.5e-27;
Matches 85; Conservative 65; Mismatches 135; Indels 68; Gaps 5;

QY 147 FTCTGHKQWVLCVSWAPDGKRLASGCKAGSIIIWDPETGQKGRPLSGHKKHINCLAWEP 206
Db :||:| | | | | :||:| ||| :| | | | :||:| | | :||:| | |
101 YALSGHRSPVTRVIFHPVFSVMVSASEDATIKVWDYETGDFE-RTLKGHTDSVQDISF-- 157

QY 207 YHRDPECRKLASAGDGCRIWVVKLGQCLMNIAGHTNAVTAVR-WGGAGLIYTSSKORT 265
Db | :||| | | | :||:| | :||:| | | :||:| | | :||:| | |
158 ---DHSGKLLASCSADMTIKLWDFQGFECIRTMHGHDNVSSVAIMPNGDHIYSASRDKT 214

QY 266 VKMWRADGILCRTFSGHAHVNNIALSTDYVLRTPGPHVPVKDRSKSHLSLSTEELQESA 325
Db :||| | :||:| | | :| | | :| | | :| | | :| | |
215 IKMWEVQTGYCVKFTFGHREWVRMVRPNQDGTL----- 247

QY 326 LKRYQAVCPDEVESILVSCDDNTLYLWRNNQKCVERTGHQNVNDVKYSPDYK----- 380
Db :||:| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
248 -----IASCSNDQTVRVVVVATKECKAELREHEHVVECISSWAPESYSSIS 293

QY 381 -----LIASAFDKSVRLWRASDGOYMATFRGHVQAVYTVAWSADSRLLIV 425
Db :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
294 EATGSETKKSGKPGPFLLSGSRDKTIKMWDVSTGMCLMTLVGHDNVRGVLFHSGGKFIL 353

QY 426 SGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKWL 478
Db | :| | | | :| :| | | :| :| :| :| :| :| :| :| :| :| :| :| :|
354 SCADDKTLRVWDYKNKRCMKTLNAHEHFTSLDFHKHTAPYVVTGSDVDTVKVW 406

RESULT 9
US-08-283-917-9
; Sequence 9, Application US/08283917
; Patent No. 5849557
; GENERAL INFORMATION:
; APPLICANT: ADACHI, HIDEKI
; APPLICANT: TSUJIMOTO, MASAFUMI
; APPLICANT: INOUE, KEIZO
; APPLICANT: ARAI, HIROYUKI
; TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME
; TITLE OF INVENTION: AND GENE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &
; ADDRESSEE: NEUSTADT,P.C.
```

STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/283,917  
FILING DATE: 03-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 209943/1993  
FILING DATE: 03-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5849557man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2292-030-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 410 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-283-917-9

Query Match 13.8%; Score 354; DB 2; Length 410;  
Best Local Similarity 24.1%; Pred. No. 4.5e-27;  
Matches 85; Conservative 65; Mismatches 135; Indels 68; Gaps 5;  
QY 147 FTCTGHKQWVLCVSWAPDGKRLASGCKAGSIIIWDPETGQKGRPLSGHKKHINCLAWEP 206  
Db 102 YALSGHRSPVTRVIFHPVFSVMVSASEDATIKVWDYETGDFE-RTLKGHTDSVQDISF-- 158  
QY 207 YHRDPECKRLASASGDGCRIVDWKLGQCLMNIAGHTNAVTAVR-WGGAGLIYTSSKDRT 265  
Db 159 ---DHSGKLLASCADMTIKLWDFQGFECIRTMHGHHDHNVSSVAIMPNGDHIIVSASRDKT 215  
QY 266 VKMRAADGILCRTFSGHAHWNNIALSTDYVLRGTFPHFPVKDRSKSHLSLSTEELQESA 325  
Db 216 IKMWEVQTGYCVKTFTHGREWVRVVRPNQDGTL----- 248  
QY 326 LKRYQAVCPDEVESLVSCDDNTLYLWNNQNKCVERTMTHQNVVNDVKYSPDVK----- 380  
Db 249 -----IASCSNDQTVRVVWVATKECKAELREHEHVVEICISWAPESSYSSIS 294  
QY 381 -----LIASASFDKSVRLWRASDGOYMATFRGHVQAVYTVANSADSRLLIV 425  
Db 295 EATGSETKKSGKPGPFLLSGSRDKTIKMDVSTGMCIMTLVGHDNWVRGVLFHSGGKFIL 354  
QY 426 SGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKWL 478  
Db 355 SCADDKTLRVWDYKKNKRCMKTLNAHEHFTVSLDFHKTAPYVVTGSDVQTVKVM 407

RESULT 10  
US-08-961-716-9  
; Sequence 9, Application US/08961716  
; Patent No. 5880272  
; GENERAL INFORMATION:  
; APPLICANT: ADACHI, HIDEKI  
; APPLICANT: TSUJIMOTO, MASAFUMI  
; APPLICANT: INOUE, KEIZO  
; APPLICANT: ARAI, HIROYUKI  
; TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME  
; TITLE OF INVENTION: AND GENE THEREOF

NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &  
ADDRESSEE: NEUSTADT,P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,716  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/283,917  
FILING DATE: 03-AUG-1994  
APPLICATION NUMBER: JP 209943/1993  
FILING DATE: 03-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5880272man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2292-030-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 410 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-716-9

Query Match 13.8%; Score 354; DB 2; Length 410;  
Best Local Similarity 24.1%; Pred. No. 4.5e-27;  
Matches 85; Conservative 65; Mismatches 135; Indels 68; Gaps 5;  
QY 147 FTCTGHKQWVLCVSWAPDGKRLASGCKAGSIIIWDPETGQKGRPLSGHKKHINCLAWEP 206  
Db 102 YALSGHRSPVTRVIFHPVFSVMVSASEDATIKVWDYETGDFE-RTLKGHTDSVQDISF-- 158  
QY 207 YHRDPECKRLASASGDGCRIVDWKLGQCLMNIAGHTNAVTAVR-WGGAGLIYTSSKDRT 265  
Db 159 ---DHSGKLLASCADMTIKLWDFQGFECIRTMHGHHDHNVSSVAIMPNGDHIIVSASRDKT 215  
QY 266 VKMRAADGILCRTFSGHAHWNNIALSTDYVLRGTFPHFPVKDRSKSHLSLSTEELQESA 325  
Db 216 IKMWEVQTGYCVKTFTHGREWVRVVRPNQDGTL----- 248  
QY 326 LKRYQAVCPDEVESLVSCDDNTLYLWNNQNKCVERTMTHQNVVNDVKYSPDVK----- 380  
Db 249 -----IASCSNDQTVRVVWVATKECKAELREHEHVVEICISWAPESSYSSIS 294  
QY 381 -----LIASASFDKSVRLWRASDGOYMATFRGHVQAVYTVANSADSRLLIV 425  
Db 295 EATGSETKKSGKPGPFLLSGSRDKTIKMDVSTGMCIMTLVGHDNWVRGVLFHSGGKFIL 354  
QY 426 SGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKWL 478  
Db 355 SCADDKTLRVWDYKKNKRCMKTLNAHEHFTVSLDFHKTAPYVVTGSDVQTVKVM 407

RESULT 11  
US-08-190-802A-51  
; Sequence 51, Application US/08190802A  
; Patent No. 5519003  
; GENERAL INFORMATION:







RESULT 13

US-08-473-089-51  
; Sequence 51, Application US/08473089  
; Patent No. 6342368  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/473,089  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2550-0025.22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 409 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: LIS1 (human), Fig. 34  
US-08-473-089-51

Query Match 13.2%; Score 340.5; DB 4; Length 409;  
Best Local Similarity 24.1%; Pred. No. 1e-25;  
Matches 85; Conservative 64; Mismatches 135; Indels 69; Gaps 6;  
QY 147 FTCTGHKQWVLCVSWAPDGKRLASGCKAGSIIIWDPETGQKGRPLSGHKHHINCLAWEP 206  
; :||: | | | | | : : :||| | | | : | | | : : :  
Db 102 YALSGHRSPVTRVIFHPVFSVMVSASEDATIKVWDYETGDFE-RTLKGHTDSVQDISF-- 158  
QY 207 YHRDPECRKLASASGDGDCRIWDVKLGQCLMNIAGHTNAVAVR-WGGAGLIYTSSKDRT 265  
; :||| | | | | :||: | | | :||: | | | | | :||: | | |  
Db 159 ---DHSGKLLASCSADMTIKLWDFQGFECIRTMGHGHDHNVSSVAIMPNGDHIVSASRDKT 215  
QY 266 VKMWRADGILCRTFSGHAHWVNNIALSTDYVLRTPFPFHPVKDRSKSHLSLSTEELOESA 325  
;||| | | | | | :||: | | | | : | | |  
Db 216 IKMWEVQTGYCVKTFTHGREWVRMVRPNQDGTL----- 248  
QY 326 LKRYQAVCPDEVESLVSCSDNTLYLWRNNQKCVERTMTHQNVVNDVKYSPDVK----- 380  
; |||: | | : | | : | | : | | : | | : | | : | | : | | :  
Db 249 -----IASCSNDQTVRVVWVAVTAECKAELREHEHVVECISSWAPESYSSIS 294  
QY 381 -----LIASAFDKSVRLWRASDGQYMATFRGHVQAVYTVAWSADSLIV 425  
; | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 295 EATGSETKKSGKPGPFLLSGSRDKT-KMWDVSTGMCLMTLVGHDNWNVRGVLFHSGGKFIL 353  
QY 426 SGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKLM 478  
; | : | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 354 SCADDKTLRVWDYKKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSDVQTVKVM 406

RESULT 14

US-08-487-072A-51  
; Sequence 51, Application US/08487072A  
; Patent No. 6423684  
; GENERAL INFORMATION:  
; APPLICANT: Mochly-Rosen, Daria  
; APPLICANT: Ron, Dorit  
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 265  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morrison & Foerster  
; STREET: 2000 Pennsylvania Avenue, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,072A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 2550-0025.20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; INFORMATION FOR SEQ ID NO: 51:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 409 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; INDIVIDUAL ISOLATE: LIS1 (human), Fig. 34  
US-08-487-072A-51

Query Match 13.2%; Score 340.5; DB 4; Length 409;  
Best Local Similarity 24.1%; Pred. No. 1e-25;  
Matches 85; Conservative 64; Mismatches 135; Indels 69; Gaps 6;  
QY 147 FTCTGHKQWVLCVSWAPDGKRLASGCKAGSIIIWDPETGQKGRPLSGHKHHINCLAWEP 206  
; :||: | | | | | : : :||| | | | : | | | : : :  
Db 102 YALSGHRSPVTRVIFHPVFSVMVSASEDATIKVWDYETGDFE-RTLKGHTDSVQDISF-- 158  
QY 207 YHRDPECRKLASASGDGDCRIWDVKLGQCLMNIAGHTNAVAVR-WGGAGLIYTSSKDRT 265  
; :||| | | | | :||: | | | :||: | | | | | :||: | | |  
Db 159 ---DHSGKLLASCSADMTIKLWDFQGFECIRTMGHGHDHNVSSVAIMPNGDHIVSASRDKT 215  
QY 266 VKMWRADGILCRTFSGHAHWVNNIALSTDYVLRTPFPFHPVKDRSKSHLSLSTEELOESA 325  
;||| | | | | | :||: | | | | : | | |  
Db 216 IKMWEVQTGYCVKTFTHGREWVRMVRPNQDGTL----- 248  
QY 326 LKRYQAVCPDEVESLVSCSDNTLYLWRNNQKCVERTMTHQNVVNDVKYSPDVK----- 380  
; |||: | | : | | : | | : | | : | | : | | : | | : | | :  
Db 249 -----IASCSNDQTVRVVWVAVTAECKAELREHEHVVECISSWAPESYSSIS 294  
QY 381 -----LIASAFDKSVRLWRASDGQYMATFRGHVQAVYTVAWSADSLIV 425  
; | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 295 EATGSETKKSGKPGPFLLSGSRDKT-KMWDVSTGMCLMTLVGHDNWNVRGVLFHSGGKFIL 353  
QY 426 SGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKLM 478  
; | : | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 354 SCADDKTLRVWDYKKNKRCMKTLNAHEHFVTSLDFHKTAPYVVTGSDVQTVKVM 406





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:59:40 ; Search time 264 Seconds  
(without alignments)  
184.043 Million cell updates/sec

Title: US-09-830-980-1  
Perfect score: 2574  
Sequence: 1 MQETDTEQATPHTIQARLV.....PDGSRVASGGKDKVIKLWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 101223694 residues 383519

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1521	59.1	485	9 US-10-132-744A-6	Sequence 6, Appli
2	1452	56.4	484	9 US-10-132-744A-2	Sequence 2, Appli
3	819	31.8	261	9 US-10-132-744A-4	Sequence 4, Appli
4	581	22.6	1356	9 US-10-077-111-10	Sequence 10, Appli
5	409.5	15.9	742	9 US-10-077-111-11	Sequence 11, Appli
6	348	13.5	521	10 US-09-764-853-449	Sequence 449, App
7	346	13.4	678	10 US-09-801-368-314	Sequence 314, App
8	344	13.4	159	9 US-09-774-639-247	Sequence 247, App
9	344	13.4	159	9 US-09-969-730-345	Sequence 345, App
10	323.5	12.6	343	9 US-10-119-932-5	Sequence 5, Appli
11	320	12.4	423	10 US-09-729-674-160	Sequence 160, App
12	313.5	12.2	569	9 US-10-038-010-8	Sequence 8, Appli
13	313.5	12.2	569	12 US-10-042-417-2	Sequence 2, Appli
14	312	12.1	1194	10 US-09-876-667-2	Sequence 2, Appli
15	312	12.1	1205	10 US-09-876-667-16	Sequence 16, Appli
16	310.5	12.1	640	9 US-10-060-019-30	Sequence 30, Appli
17	308	12.0	375	9 US-10-119-932-1	Sequence 1, Appli
18	302	11.7	540	9 US-09-213-888-7	Sequence 7, Appli
19	302	11.7	540	9 US-09-213-888-10	Sequence 10, Appli

20	302	11.7	540	9 US-09-328-877A-7	Sequence 7, Appli
21	302	11.7	540	9 US-09-328-877A-10	Sequence 10, Appli
22	302	11.7	545	9 US-09-213-888-6	Sequence 6, Appli
23	302	11.7	545	9 US-09-328-877A-6	Sequence 6, Appli
24	302	11.7	553	9 US-09-213-888-5	Sequence 5, Appli
25	302	11.7	553	9 US-09-328-877A-5	Sequence 5, Appli
26	302	11.7	559	9 US-09-213-888-9	Sequence 9, Appli
27	302	11.7	559	9 US-09-328-877A-9	Sequence 9, Appli
28	302	11.7	589	9 US-09-213-888-8	Sequence 8, Appli
29	302	11.7	589	9 US-09-328-877A-8	Sequence 8, Appli
30	302	11.7	592	9 US-09-213-888-4	Sequence 4, Appli
31	302	11.7	592	9 US-09-328-877A-4	Sequence 4, Appli
32	302	11.7	626	9 US-09-213-888-21	Sequence 21, Appli
33	302	11.7	626	9 US-09-328-877A-21	Sequence 21, Appli
34	302	11.7	627	9 US-09-213-888-3	Sequence 3, Appli
35	302	11.7	627	9 US-09-328-877A-3	Sequence 3, Appli
36	302	11.7	666	9 US-09-213-888-27	Sequence 27, Appli
37	302	11.7	666	9 US-09-328-877A-27	Sequence 27, Appli
38	302	11.7	669	9 US-09-213-888-25	Sequence 25, Appli
39	302	11.7	669	9 US-09-328-877A-25	Sequence 25, Appli
40	299	11.6	613	9 US-10-108-605-181	Sequence 181, App
41	294.5	11.4	713	10 US-09-801-368-408	Sequence 408, App
42	292.5	11.4	402	10 US-09-925-300-1320	Sequence 1320, Ap
43	292.5	11.4	404	9 US-09-943-689A-2	Sequence 2, Appli
44	289.5	11.2	400	9 US-09-943-689A-4	Sequence 4, Appli
45	288.5	11.2	404	10 US-09-908-805B-21	Sequence 21, Appli

ALIGNMENTS

RESULT 1  
US-10-132-744A-6  
; Sequence 6, Application US/10132744A  
; Publication No. US20030027261A1  
; GENERAL INFORMATION:  
; APPLICANT: Utku, Nalan  
; TITLE OF INVENTION: No. US20030027261A1 genes Tzap7/A, Tzap7/B and Tzap7 invoi  
; TITLE OF INVENTION: activation and uses thereof  
; FILE REFERENCE: Utku-4 CON  
; CURRENT APPLICATION NUMBER: US/10/132,744A  
; PRIOR FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: PCT/EP00/10670  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 60/185,016  
; PRIOR FILING DATE: 2000-02-25  
; PRIOR APPLICATION NUMBER: 60/162,675  
; PRIOR FILING DATE: 1999-11-01  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: miscellaneous feature  
; LOCATION: 379  
; OTHER INFORMATION: variable amino acid  
US-10-132-744A-6

Query Match	59.1%;	Score 1521;	DB 9;	Length 485;
Best Local Similarity	58.6%;	Pred. No. 2.7e-123;		
Matches	279;	Conservative	73;	Mismatches 118;
			Indels	6;
			Gaps	3;
QY	9	EATPHTIQARLVYTGEEA----	GPPIDLPAGITITQOLGLICNALLKNEEATPYLFFVGED	64
Db	8	EAVARDVQRLVQFQDEGGQLLGSFPDVPVDITPRLQLVCNALLAQEDPLPLAFFVHDA	67	
QY	65	EIKKSLDTLDLASVDTENVIDIVYQAVKVRVTRCTSSMPGHAEAVVSLNFSPDGA	124	
Db	68	EIVSSLGKLTLESQAVETEKVLDIIYQPAIFRVRVTRCTSSLEGHSEAVISVAFSPGK	127	
QY	125	HLASGSGDTTVRLWDLNLTETPHFTCTGCHKQWLVCSWAPDCKRLASGCKAGSIWDPET	184	

Db	128	YLAGSGD	T	T	V	R	F	W	L	D	S	T	E	P	H	E	T	C	K	G	H	R	H	W	L	S	I	S	W	S	P	D	G	K	L	A	S	C	K	N	G	Q	I	L	L	W	D	P	S	T	187									
Qy	185	GQQGR	P	L	S	G	H	K	K	H	I	N	C	L	A	W	E	P	Y	R	D	P	E	C	R	K	L	A	S	A	S	G	D	G	C	R	I	W	D	V	K	L	G	Q	I	L	M	N	I	A	G	H	T	N	244					
Db	188	GKQV	C	R	T	L	A	G	S	K	W	I	T	G	L	S	W	E	P	L	H	A	N	P	E	C	R	Y	V	A	S	S	K	D	G	S	V	R	I	W	D	T	T	A	G	C	E	R	I	L	T	G	H	Q	247					
Qy	245	AVTAV	R	W	G	G	A	G	L	I	T	S	S	K	D	R	T	V	M	W	R	A	A	D	G	I	L	C	R	T	F	S	G	H	A	H	V	N	N	I	A	L	S	T	D	V	L	R	T	G	P	F	H	304						
Db	248	SVTC	L	R	W	G	G	D	L	L	Y	S	A	S	Q	D	R	T	I	K	V	W	R	A	H	D	G	V	L	C	R	T	L	Q	H	G	H	W	V	N	T	M	A	L	S	T	D	Y	A	L	R	T	G	A	F	E	307			
Qy	305	PV	K	-	D	R	S	K	S	H	L	S	T	E	E	L	Q	E	S	A	L	K	R	Y	Q	A	V	C	P	D	E	V	E	S	L	V	S	C	D	D	N	T	L	Y	L	M	R	N	Q	N	-	K	V	E	R	362				
Db	308	P	A	E	A	S	V	N	P	Q	D	L	Q	G	S	L	Q	E	L	K	E	R	A	L	S	R	N	L	V	R	G	O	G	P	E	R	L	V	S	G	S	D	F	T	L	F	L	W	S	P	A	E	D	K	K	P	L	T	R	367
Qy	363	MTG	H	Q	N	V	V	N	D	V	K	Y	S	P	D	V	K	L	I	A	S	F	D	K	S	V	R	L	W	R	A	S	D	Q	Y	M	A	T	E	R	G	H	V	Q	A	V	T	V	A	W	S	A	D	S	R	422				
Db	368	MTG	H	Q	A	L	I	N	Q	V	X	F	S	P	D	S	R	I	V	A	S	A	S	F	D	K	S	I	K	L	W	D	G	R	T	G	K	Y	L	A	S	L	R	G	H	V	A	A	V	Y	Q	I	A	W	S	A	D	S	R	427
Qy	423	L	I	V	S	G	K	D	S	T	L	K	V	W	S	V	Q	T	K	L	A	Q	E	L	P	G	H	A	D	E	V	F	G	V	D	W	A	P	D	G	S	R	V	A	S	G	G	K	D	K	V	I	K	L	W	478				
Db	428	L	L	V	S	G	S	D	S	T	L	K	V	W	D	V	K	A	Q	K	L	A	M	D	L	P	G	H	A	D	E	V	A	D	W	S	P	D	G	O	R	V	A	S	G	G	K	D	C	L	R	I	W	483						

```

RESULT 2
US-10-132-744A-2
; Sequence 2, Application US/10132744A
; Publication No. US20030027261A1
; GENERAL INFORMATION:
; APPLICANT: Utku, Nalan
; TITLE OF INVENTION: No. US20030027261A1el genes Tzap7/A, Tzap7/B and Tzap7 involved in
; TITLE OF INVENTION: activation and uses thereof
; FILE REFERENCE: Utku-4 CON
; CURRENT APPLICATION NUMBER: US/10/132,744A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: PCT/EP00/10670
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/185,016
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 60/162,675
; PRIOR FILING DATE: 1999-11-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 2
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-132-744A-2

```

QY	305	PVK-DRSKSHLSLSTEELEQESALKRYOAVCPDEVESLVSCDDNTLYLWRNNQN-KCVER	362
Dd	307	PAEASVNPQDLQSGSLQELKERALSRYNLVRGQGPRLVSGSDDFTFLWSPAEDKKPLTR	366
QY	363	MTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMATFRGHVQAVYTVAWSADSR	422
Dd	367	MTGHQALINQVLFSPPDSRIVASAFDKSIKLWDGRTGKYLASLRGHVAAVYQIAWSADSR	426
QY	423	LIVSGSKDSTLKVWSVQTCKLAQLPQHADADEVFGVDWAPDGSRVASGGKDVKIKLW	478
Dd	427	LIVSGSSDSTLKVWDVKAQKCLAMDLPQHADADEVYAVDWSPDGORVASGGKDKCLRIR	482
 RESULT 3 US-10-132-744A-4 ; Sequence 4, Application US/10132744A ; Publication No. US20030027261A1 ; GENERAL INFORMATION: ; APPLICANT: Utku, Nalan ; TITLE OF INVENTION: No. US20030027261A1 genes Tzap7/A, Tzap7/B and Tzap7 inv ; TITLE OF INVENTION: activation and uses thereof ; FILE REFERENCE: Utku-4 CON ; CURRENT APPLICATION NUMBER: US/10/132,744A ; CURRENT FILING DATE: 2002-04-24 ; PRIOR APPLICATION NUMBER: PCT/EP00/10670 ; PRIOR FILING DATE: 2000-10-30 ; PRIOR APPLICATION NUMBER: 60/185,016 ; PRIOR FILING DATE: 2000-02-25 ; PRIOR APPLICATION NUMBER: 60/162,675 ; PRIOR FILING DATE: 1999-11-01 ; NUMBER OF SEQ ID NOS: 6 ; SOFTWARE: PatentIn Ver. 2.1 ; SEQ ID NO 4 ; LENGTH: 261 ; TYPE: PRT ; ORGANISM: Homo sapiens US-10-132-744A-4			
 Query Match            31.8%;     Score 819;     DB 9;     Length 261; Best Local Similarity   58.8%;     Pred. No. 6.3e-63; Matches 150;     Conservative   41;     Mismatches   62;     Indels       2;     Gaps       2;			
QY	226	RIWDVKLGQCLMNIAGHTNAVTAVRWGGAGLIYTSKSDRTVKMWRAADGILCRTESGHAH	285
Dd	5	RIWDTTAGRCERILTGHQTQSVTCLRWGGDGLLYSASQDRTIKVWRAHDGVLCRTLQHGHH	64
QY	286	WVNIALSTDYVLTGTFHPVK-DRSKSHLSLSTEELQEESALKRYQAACVPDESLSVCS	344
Dd	65	WVNTMALSTDYALRTGAFFPAEASVNPQDLQSGSQELKERALSRYNLVRGQGPRLVSGS	124
QY	345	DDNTLYLWRNNQN-KCVERMTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDGQYMA	403
Dd	125	DDFTFLWSPAEDKKPLTRMTGHQALINQVLFSPPDSRIVASAFDKSIKLWDGRTGKYLA	184
QY	404	TFRGHVQAVYTVAWSADSRLLIVSGSKDSLTKVWSVQTKKLAQELPGHADEVFEGVDWAPDG	463
Dd	185	SLRGHVAAVYQIAWSADSRLLVSGSSDSTLKVWDVKAQKCLAMDLPQHADADEVYAVDWSPDG	244
QY	464	SRVASGGKDVKIKLW	478
Dd	245	ORVASGGKDKCLRIR	259

RESULT 4  
US-10-077-111-10  
; Sequence 10, Application US/10077111  
; Publication No. US20020187492A1  
; GENERAL INFORMATION:  
; APPLICANT: Todderud, C. Gordon  
; APPLICANT: Finger, Joshua N.  
; APPLICANT: Rillema, Jill  
; TITLE OF INVENTION: TBA  
; FILE REFERENCE: 3053-4114US2



; CURRENT APPLICATION NUMBER: US/10/077,111  
; CURRENT FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/294,181  
; PRIOR FILING DATE: 2001-05-29  
; PRIOR APPLICATION NUMBER: 60/269,366  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 1356  
; TYPE: PRT  
; ORGANISM: Podospora anserina  
; FEATURE:  
; OTHER INFORMATION: beta transducin-like protein encoded by the  
; OTHER INFORMATION: het-e-1 gene  
US-10-077-111-10

Query Match 22.6%; Score 581; DB 9; Length 1356;  
Best Local Similarity 27.2%; Pred. No. 2.5e-41;  
Matches 138; Conservative 81; Mismatches 182; Indels 106; Gaps 7;  
US-10-077-111-10

QY 13 HTIQARLVYTGEGAPPIDLPAGITTTQQLGLICNALKNEEATPY---LFFVGEDEIKKS 69  
Db 763 HTIRGLIAIVRDG-----YRFALSYRMILIEKAPLQAYTSALVFAPRDSMIKK 810  
QY 70 LEDTLDLASVDTENVIDIVYQPAVKVRPVTRCTSSMPGHAEAVVSLNFSPPGHAHLASG 129  
Db 811 IFKKEPGWISTISVVEAEW-----NACTQTLEGGSSVLSVAFSADQORVASG 859  
QY 130 SGDTVRLWDLNTEPHTCTCHKQWVLCVSWAPDGRKRLASGCKAGSIIIWDPETGQQKG 189  
Db 860 SDDKTIKIWDTASGTGTQTLEGGGVSWSVAFSPDRERVASGSDDKTIKIWDASGTCT- 918  
QY 190 RPLSGHKKHINCLAWEPYHR-----DPE 212  
Db 919 QTLEGGGRVQSVAFSPDQQRVASGSDDHITIKIWDASGTCTQTLEGGSSVLSVAFSPD 978  
QY 213 CRKLASAGDGCRIWDVKGQCLMNIAGHTNAVAVRWGGAG-LIYTSSKDRTVKMWRA 271  
Db 979 QORVASGSGDKTIKIWDTASGTCTQTLEGGGVSWSVAFSPDQQRVASGSDDKTIKIWDT 1038  
QY 272 ADGILCRTFSGHAHWNNIALSTDIVLRGPFHPVKDRSKSHLSLSTEELQESALKRYQA 331  
Db 1039 ASGTCTQTLEGGGWVQSVFSPD----- 1062  
QY 332 VCPDEVESLVSCSDNTLYLWRNNQNKVERMTGHQNVVNDVKYSPDVKLIASAFDKSV 391  
Db 1063 -----QORVASGSDDHITIKIWDASVGTCTQTLEGGGVSWSVAFSPDQORVASGSDGTI 1117  
QY 392 RLWRASDQGYMATFRGHVQAVYTVAWSADSRLLIVSGSKDSTLKVWSVQTKKLAQELPGHA 451  
Db 1118 KIWDASGTCTQTLEGGGWSVAFSPDQORVASGSDGTIKIWDASGTCTQTLEGGH 1177  
QY 452 DEVFGVDWAPDGSRVASGGKDKVIKLW 478  
Db 1178 GWVQSVAFSPDQORVASGSDDKTIKI 1204

RESULT 5  
US-10-077-111-11  
; Sequence 11, Application US/10077111  
; Publication No. US20020187492A1  
; GENERAL INFORMATION:  
; APPLICANT: Todderud, C. Gordon  
; APPLICANT: Finger, Joshua N.  
; APPLICANT: Rillema, Jill  
; TITLE OF INVENTION: TBA  
; FILE REFERENCE: 3053-4114US2  
; CURRENT APPLICATION NUMBER: US/10/077,111  
; CURRENT FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: 60/294,181  
; PRIOR FILING DATE: 2001-05-29  
; PRIOR APPLICATION NUMBER: 60/269,366

; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 11  
; LENGTH: 742  
; TYPE: PRT  
; ORGANISM: Thermomonospora curvata  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence encoded by the PKWA gene  
US-10-077-111-11

Query Match 15.9%; Score 409.5; DB 9; Length 742;  
Best Local Similarity 27.4%; Pred. No. 7.4e-27;  
Matches 100; Conservative 61; Mismatches 113; Indels 91; Gaps 8;  
US-10-077-111-11

QY 115 VSLNFSPPDGAHLASGSDTTVRLWDLNTEPHTCTCHKQWVLCVSWAPDGRKLASGCKA 174  
Db 463 VAVAFSPGGSLAGSGDKLIHVWDVASGDELHTLEGHTDWVRAVAFSPDGALLASGSD 522  
QY 175 GSIIWDPETGQOKGRPLSGHKKHINCLAWEPYHRDPECKRLASASGDGDCRIWDVKLGQ 234  
Db 523 ATVRLWDVAAAEEERA-VFEGHTHYVLDIAF-----SPDGSMVASGSRDGTARLWNVATGT 576  
QY 235 CLMNIAGHTNAVAVRWGGAG-LIYTSSKDRTVKMWRAADGILCRTFSGHAHWNNIALS 293  
Db 577 ERAVLKGHTDYVYAVAFSPDGSMVASGSRDGTIRLWDVATG----- 617  
QY 294 TDYVLRGTPHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDNTLYLWR 353  
Db 618 -----KER-----DVLQAPA----- 627  
QY 354 NNQNKCVERTGHQNVVNDVKYSPDVKLIASAFDKSVRLWRASDQGYMATFRGHVQAVY 413  
Db 628 -----ENVVS-LAFSPDGSMLVHGS-DSTVHLWDVASGEALHTFEHTDWR 672  
QY 414 TVAWSADSRLLIVSGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDK 473  
Db 673 AVAFSPDGALLASGSDRTIRLWDVAAQEEHTTLEGHTPEVHSAFHPEGTTLASASEDG 732  
QY 474 VIKLW 478  
Db 733 TIRIW 737

RESULT 6  
US-09-764-853-449  
; Sequence 449, Application US/09764853  
; Patent No. US20020090672A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJZ06  
; CURRENT APPLICATION NUMBER: US/09/764,853  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 939  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 449  
; LENGTH: 521  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-853-449

Query Match 13.5%; Score 348; DB 10; Length 521;  
Best Local Similarity 23.2%; Pred. No. 9.4e-22;  
Matches 86; Conservative 56; Mismatches 139; Indels 90; Gaps 5;  
US-09-764-853-449

QY 109 GHAEAVVSLNFSPPDGAHLASGSDTTVRLWDLNTEPHTCTCHKQWVLCVSWAPDGRKL 168  
Db 59 GHKAAITSLDSPNGKQLATASDWTFLMLWNFKPHARAYRYVGHKDVVTSVQFSPHGNLL 118  
QY 169 ASGCKAGSIIWDPETGQOKGRPLSGHKKHINCLAWEPYHRDPECKRLASASGDGDCRIW 228

Db 119 ASASRDRTVRLWIPD---KRGK----- 137

QY 229 DVKLGQCLMNIAGHTNAVAVRWGGAG-LIYTSSKDRTVKMWRAADGILCRTLFSGHAHW 287

Db 138 -----FSEFKAHTAPVRSVDFESADGQFLATASEDKSIKVMWSYRQRFYLSYLRHTHW 190

QY 288 NNIALSTDYVLRTPGFHPVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDDN 347

Db 191 -----RCAKFS-----DGRLLVSCSEDK 209

QY 348 TLYLWRNNQKCVERTGHQNVVNDVKYSPDVKLIASASFDKSVRLWRASDQYMATFRG 407

Db 210 TIKIWDTTNKQCVNNFSDSVGFANFVDFNPSGTCIASAGSDQTVKVVWDVVRVKNLLQHYQV 269

QY 408 HVQAVYTVAWSADSRLLIVSGSKDSTLKWVSVQTKLAQELPCHADEVFGVDWAPDGSRVA 467

Db 270 HSGGVNCISFHPSPGNYLITASSDGTILKLDLLEGRLIYTLQHTGVPVFTVSFSGGELFA 329

QY 468 SGGKDKVIKLV 478

Db 330 SGGADTQVLLW 340

RESULT 7

US-09-801-368-314

; Sequence 314, Application US/09801368

; Patent No. US20020128250A1

; GENERAL INFORMATION:

; APPLICANT: Busby, Robert

; APPLICANT: Cali, Brian

; APPLICANT: Hecht, Peter

; APPLICANT: Holtzman, Doug

; APPLICANT: Madden, Kevin

; APPLICANT: Maxon, Mary

; APPLICANT: Milne, Todd

; APPLICANT: No. US20020128250A1man, Thea

; APPLICANT: Royer, John

; APPLICANT: Salama, Sofie

; APPLICANT: Sherman, Amir

; APPLICANT: Silva, Jeff

; APPLICANT: Summers, Eric

; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi

; FILE REFERENCE: 109272.147

; CURRENT APPLICATION NUMBER: US/09/801.368

; CURRENT FILING DATE: 2001-03-07

; PRIOR APPLICATION NUMBER: US 09/487,558

; PRIOR FILING DATE: 2000-01-19

; PRIOR APPLICATION NUMBER: US 60/160,587

; PRIOR FILING DATE: 1999-10-20

; NUMBER OF SEQ ID NOS: 440

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 314

; LENGTH: 678

; TYPE: PRT

; ORGANISM: Aspergillus nidulans

US-09-801-368-314

Query Match 13.4%; Score 346; DB 10; Length 678;

Best Local Similarity 24.2%; Pred. No. 2e-21;

Matches 95; Conservative 60; Mismatches 118; Indels 120; Gaps 11;

QY 148 TCTGKHQWVLCVSWAPDGRRLASGCKAGSIIIDPETGQOKGRPLSGHKKHINCLAWEPY 207

Db 344 TFKGHTNGVMCLQF--EDNIATGSDYDTIKIWDTEGEEL-RTLGHESGIRCLQFDD- 399

QY 208 HRDPECKLASAGDGDCCIWDVKLGQCLMNIAGHTNAVTAVRWGGAGLIYTSKDRTVK 267

Db 400 -----TKLISGMDRTIKVNNWRTGECISTYTGHRGGVIGLHF-DASILASGVDKTVK 452

QY 268 MWRAADGILCRTFS--GHAHVNNIALSTDYVLRTPGFHPVKDRSKSHLSLSTEELQESA 325

Db 453 TWNFEDK---STFSLRGHTDWNVAVRVDFT----- 478

QY 326 LKRYQAVCPDEVESLYSCSDDNTLYLWRNNQKCVERTMTGHQNVVNDV----- 373

Db 479 -----SSRTVFSASDDCTVRLWDLDTKTCIRTFHGHVGVQVQVPLPREFEFEH 528

QY 374 -----KYSPDVKLIASASF 387

Db 529 DAECENDDLSTTSGDANPPSIQASMGLEPNAAYSQSSAFGTSDNGRAAPRYMTSA-L 587

QY 388 DKSVRRLWRASDQYMATFRGHVQAVYTVAWSADSRLLIVSGSKDSTLKWVSVQTKKLAQEL 447

Db 588 DSTIRLWETTGRCLRTFFGHLEGVW--ALGADTLRIVSGAEDRMIKIWDPTGKCERTF 645

QY 448 PGHADEVFGVDWAPDGSRVASGGKDKVIKLVAY 480

Db 646 TGHSGPVTICIGLG--DSRFATGSEDCVVRMYSF 676

RESULT 8

US-09-774-639-247

; Sequence 247, Application US/09774639

; Publication No. US20030003555A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: 90 Human Secreted Proteins

; FILE REFERENCE: PZ013P1

; CURRENT APPLICATION NUMBER: US/09/774,639

; CURRENT FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/244,112

; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-04

; NUMBER OF SEQ ID NOS: 371

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 247

; LENGTH: 159

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (63)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (137)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-774-639-247

Query Match 13.4%; Score 344; DB 9; Length 159;

Best Local Similarity 53.3%; Pred. No. 3.9e-22;

Matches 72; Conservative 20; Mismatches 41; Indels 2; Gaps 2;

QY 296 YVLRTPGFHPVK-DRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDDNTLYLWRN 354

Db 1 YALRTGAFEPAEASVNPQDLQGLSLQELKERALSRYNLVRGQGPRLVSGSDDFTLFLWSP 60

QY 355 NQ-NKCVERTGHQNVVNDVKYSPDVKLIASASFDKSVRLWRASDQYMATFRGHVQAVY 413

Db 61 AEXKKPLTRMTGHQALINQVLFSPDSRIVASASFDKSIKLDWDRGTGKYLASLRGHVAAVY 120

QY 414 TVAWSADSRLLIVSGS 428

Db 121 QIAWSADSRLLIVSGS 135

RESULT 9

US-09-969-730-345

; Sequence 345, Application US/09969730

; Publication No. US20030054443A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et al.

; TITLE OF INVENTION: 90 Human Secreted Proteins

; FILE REFERENCE: PZ013P2

; CURRENT APPLICATION NUMBER: US/09/969,730

; CURRENT FILING DATE: 2001-10-04

; PRIOR APPLICATION NUMBER: 09/774,639

; PRIOR FILING DATE: 2001-02-01



QY 166 KRLASGCKAGSIIIWDPETGQKGRPLSGHKKHINCLAWEPYHRDPECKRLASASGDGDC 225  
Db 109 SQIVSASPDKTVRAWDVETGKI-KKMAEHSFVNSCC--PTRRGPPL--IISGSDDGTA 163  
QY 226 RIWDVKLGQCLMNIAGHTNAVTAVRWG-GAGLIYTSKDRTVKMWRAADGILCRTEFSGHA 284  
Db 164 KLWDMRQRGAIQTFFPKYO-ITAVSFDAADKIETGGVDNDVKVMDLRKGEATMTLEGHQ 222  
QY 285 HWNNIALSTD--YVLRTPGPHVPKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVS 342  
Db 223 DTITGMSLSPDGSYLLTNGM----- 242  
QY 343 CSDDNTLYLWENN---ONKCVERTGHQNVVNDVKYSPDVKLIASASFDSKSVRLWRASD 398  
Db 243 ---DNKLCVWDMRYPAPQNCVKIFEHGHQ-----NFEKN----- 274  
QY 399 GOYMATERGHVQAVYTVAMSADSRLLIVSGSKDSTLKVWSVQTKKLAQELPGHADEVFGVD 458  
Db 275 -----LLKCSWSPDGTQKVTAGSSDRMVHWDTTSRRTIYKLPGHTGSGVNECV 321  
QY 459 WAPDGSRVASGGKDKVIKL 477  
Db 322 FHPTPIIGSCSSDKNIYL 340

RESULT 11  
US-09-729-674-160  
; Sequence 160, Application US/09729674  
; Patent No. US20010039335A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fecthel, Kim  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: 6055-64X  
; CURRENT APPLICATION NUMBER: US/09/729,674  
; CURRENT FILING DATE: 2000-12-04  
; PRIOR APPLICATION NUMBER: 09/539,330  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 283  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 160  
; LENGTH: 423  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-729-674-160

Query Match 12.4%; Score 320; DB 10; Length 423;  
Best Local Similarity 23.7%; Pred. No. 1.9e-19;  
Matches 113; Conservative 79; Mismatches 175; Indels 110; Gaps 20;  
QY 29 PIDLPAGITTQGLICNALKNE---EATPYLFFVGEDEIKKSLEDTLDLASVDTEV 84  
Db 21 PFSIPAAASEIADLSNIINKLKDNFHKHVEFDLTKGQFLRMPDLKHMENISSEEV 80  
QY 85 IDIVYQPOAVEKVRPVTRCTSSMPG----HAEAVVSNFSPDGAH--LASGSDTTVRLW 138  
Db 81 VEIEY-----VEKYTAPQEQCMFHDWDWISSIK----GAEBWILTGSYDKTSRIW 126  
QY 139 DLNFTPHFTCTGCHKQWVLCVSWAPDGRKRLASGCKAGSIIIWDPETGQKGRPLSGHKKH 198

Db 127 SLEKKS-INTIVGHTDVVKDVAW-----KKDS 153  
QY 199 INCLAWEPYHRDPECKRLASASGDGDCRIWD-----VKLGQCLMNIAGHTNAVTAVRW 251  
Db 154 LSCL-----LLSASMDQTILLWENNVERNKKALHC---CRGHAGSVDSIAV 197  
QY 252 GGAGLIYTS-SKDRTVKMWRAADGILCRTEFSGHAHWNNIALSTDYVLRTPGPHVPKDRS 310  
Db 198 DSGGTKFCGSGWCKMLKIW-----STVPTDEEDEMEESTNRPRKKQK 239  
QY 311 KSHLSLSTEEL~-QESALKRYQAVCPDEVESLVCSDDDNTLYLWRNNQNKCVERTMTHQN 368  
Db 240 TEQLGLTRTPIVTLSGHMEAVSSVLWSDAEEICASWDHTIRVWDVESGSLKSTLTGNK- 298  
QY 369 VVNDVKYSPDVKLIASASFDSKSVRLW--RASDGQYMA-TFRGHVQAVYTVAWS-ADSRLLI 424  
Db 299 VENCISYSPICKRLASGSDRHIRLWDPRTKDGSLVSLTSHTGWVTSVKWSPTHEQQL 358  
QY 425 VSGSKDSTLKVWSVQT-KKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKLWAY 480  
Db 359 ISGSLDNIIVKLWDTRSCKAPLYDLAAHEDKVLSDWT-DTGLLLSGGADN--KLYSY 412

RESULT 12  
US-10-038-010-8  
; Sequence 8, Application US/10038010  
; Publication No. US20030040089A1  
; GENERAL INFORMATION:  
; APPLICANT: HYBRIGENICS  
; APPLICANT: Pierre, Legrain  
; TITLE OF INVENTION: Protein-protein interactions in adipocyte cells  
; FILE REFERENCE: B4767A  
; CURRENT APPLICATION NUMBER: US/10/038,010  
; CURRENT FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: US 60/259,377  
; PRIOR FILING DATE: 2001-01-02  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 569  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: beta-TrCpl  
; LOCATION: (1)..(569)  
; OTHER INFORMATION:  
US-10-038-010-8

Query Match 12.2%; Score 313.5; DB 9; Length 569;  
Best Local Similarity 24.6%; Pred. No. 1e-18;  
Matches 90; Conservative 59; Mismatches 132; Indels 85; Gaps 12;  
QY 139 DLNTEPHFTCTGH-----KQWVLCVSWAPDGRKRLASGCKAGSIIIWDPETGQ 187  
Db 242 DIETIESNWRRCGRHSLQRHCRSETSKGVYCIQY--DDQKIVSGLRDNTIKIWDKNTLEC 299  
QY 188 KGRPLSGHKKHINCLAWEPYHRDPECKRLASASGDGDCRIWDVKLGQCLMNIAGHTNAVT 247  
Db 300 K-RILTGTGTVLCIQYDE-----RVILTSSDSTVRVWDVNTGEMLNTLIHCEAVL 351  
QY 248 AVRWGGAGLIYTSSKDRTVKMWRAA---DGILCRTEFSGHAHWNNIALSTDYVLRTPGFH 304  
Db 352 HLFENN-GMVTCSKDRSIAVWDMASPTDITLRRVLVGHRAAVNVVDFDDKYI----- 403  
QY 305 PVKDRSKSHLSLSTEELQESALKRYQAVCPDEVESLVCSDDDNTLYLWRNNQNKCVERTM 364  
Db 404 -----VSAGSDRTTKVWNTSTCEVFTLN 427  
QY 365 GHQNVVNDVKYSPDVKLIASASFDSKSVRLWRASDGQYMATFRGHVQAVYTVANSADSRLLI 424  
Db 428 GHRGRIACLOQYRD--RLVYSGSSDNTIRLWDIECGACLRVLEGHELIVRCIRF--DNKRI 483  
QY 425 VSGSKDSTLKVWSVQTKKLAQELPG-----HADEVFGVDWAPDGSRVASGGKDKV 474









GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:54:30 ; Search time 45 Seconds  
(without alignments)  
1025.434 Million cell updates/sec

Title: US-09-830-980-1  
Perfect score: 2574  
Sequence: 1 MQETDTEQEATPHTIQARLV.....PDGSRVASGGKDKVIKLWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1318.5	51.2	473	2 T33805	hypothetical prote
2	1092	42.4	515	2 S19487	hypothetical prote
3	1047.5	40.7	502	2 T41148	trp-asp repeat con
4	599	23.3	1258	2 AI2155	WD-repeat protein
5	597	23.2	1683	2 AF2071	WD-40 repeat prote
6	590	22.9	1526	2 AC2239	WD-40 repeat prote
7	581	22.6	1356	2 T18521	beta transducin-li
8	538.5	20.9	1708	2 AE1866	WD-40 repeat prote
9	533	20.7	1227	2 AE1810	WD-40 repeat prote
10	532.5	20.7	1711	2 AD1842	WD-40 repeat prote
11	532	20.7	1747	2 AC1842	WD-40 repeat prote
12	497	19.3	934	2 AG1889	WD-40 repeat prote
13	496.5	19.3	1189	2 AI2493	WD-repeat protein
14	486	18.9	1551	2 AB2410	WD-repeat protein
15	476.5	18.5	676	2 AH2195	hypothetical prote
16	469	18.2	1189	2 AH2154	WD-repeat protein
17	451	17.5	677	2 AE1861	serine/threonine k
18	436	16.9	304	2 AG1837	WD-40 repeat prote
19	433	16.8	1693	2 S76086	beta transducin-li
20	425.5	16.5	1049	2 T42045	beta transducin-li
21	424.5	16.5	559	2 AB2202	hypothetical prote
22	410.5	15.9	317	2 T46032	WD-40 repeat regul
23	400	15.5	589	2 AG2400	WD-repeat protein
24	372	14.5	376	2 T19266	hypothetical prote
25	371	14.4	786	2 AG2375	WD-40 repeat-prote
26	368.5	14.3	777	2 T41075	hypothetical WD-re
27	365.5	14.2	876	2 T51507	WD40-repeat protei
28	361.5	14.0	501	2 T27513	hypothetical prote
29	359	13.9	614	2 S58306	WD-40 repeat regul

30	355	13.8	333	2 G85034	probable WD-repeat
31	354	13.8	409	2 S36113	LIS-1 protein - hu
32	354	13.8	410	2 S48052	platelet-activatin
33	347.5	13.5	395	2 T23317	hypothetical prote
34	344.5	13.4	265	2 AF1890	WD-repeat protein
35	327.5	12.7	357	2 AI2099	WD-40 repeat prote
36	323.5	12.6	343	2 C84870	probable splicing
37	322	12.5	342	2 AE2490	WD-repeat protein
38	318.5	12.4	586	2 T38992	WD-40 repeat regul
39	314.5	12.2	518	2 B48088	beta-transducin re
40	314.5	12.2	598	2 AE2415	WD-repeat protein
41	312.5	12.1	323	2 T02617	hypothetical prote
42	312.5	12.1	605	2 T38932	probable sulfur me
43	312.5	12.1	701	2 T16607	hypothetical prote
44	312	12.1	1194	2 T03818	apoptotic proteina
45	310.5	12.1	640	2 S49932	MET30 protein - ye

ALIGNMENTS

RESULT 1  
T33805  
hypothetical protein W07E6.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 26-May-2000  
C:Accession: T33805  
R:Latreille, P.; Wamsley, P.  
submitted to the EMBL Data Library, November 1998  
A:Description: The sequence of C. elegans cosmid W07E6.  
A:Reference number: Z21414  
A:Accession: T33805  
A:Status: preliminary; translated from CB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-473 <LAT>  
A:Cross-references: EMBL:AF106576; PIDN:AAC78176.1; GSPDB:GN00020; CESP:W07E6.2  
A:Experimental source: strain Bristol N2; clone W07E6  
C:Genetics:  
A:Gene: CESP:W07E6.2  
A:Map position: 2  
A:Introns: 46/2; 77/3; 103/1; 195/2; 256/3; 311/3; 399/2  
C:Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match	51.2%	Score 1318.5;	DB 2;	Length 473;
Best Local Similarity	53.8%	Pred. No. 9e-95;		
Matches 250;	Conservative 72;	Mismatches 122;	Indels 21;	Gaps 7;
Qy	24	EEAGPPIDLPAGITTTQGLICNALL-KNEEATPYLFFVGE-----DEIKKSLDITDLA	77	
Db	18	ELGGGILVPVDISTNELQLCNQLGSSDDPVPISFFTTGEGAEIVDSIRKSL--	72	
Qy	78	SVDTENVIDIVYQPAVFRVPVTRCTSSMPGHAENVSLNFSFDPGAHLASGSDTTVRL	137	
Db	73	-IDFETTLKLVYQPAVFRVPVTRCSASIPGHGEPVISAQFSPDGRGLASGSDQTMRI	131	
Qy	138	WDLNTETPHFTCTGHKQWVLCVSWAPDGRKRLASGCKAGSIIINWDPETQQKGRPLSGHK	197	
Db	132	WDIELEPLHTCKSHKSWVLCIAWSPDATKIASACKNGEICINWAKTGEQIGKTLKRHKQ	191	
Qy	198	HINCLAWEPYHRDPECRKLASASGDGDCRIWDVVLQCLMNIAGHTNAVTVRWGGAGLI	257	
Db	192	WITSLAWQPMHKDPTCRLILASCGKGNIFINDTVQGVVVRCLSGHTASVTCLRWGGGLI	251	
Qy	258	YTSKDRTVKMWRADGILCRTFSGHAHWNNIALSTDYVLRGTPFHPVDRKSKSHLSLS	317	
Db	252	YSGQDRTVKMWRADGVMCRNMTGHAHWINTLALNTDYALRTSCFEPKRCIKPD---T	308	
Qy	318	TEELQESALKRYQAVCPDEV---ESLVSCDDNTLYLWRNNQNK-CVERMTGHQNVNDV	373	
Db	309	VEECQKVAQTRYEAL--ETAGGERLVSGSDDFTLFWMNPKETKQINRMTHGMQLVNVQV	366	
Qy	374	KYSPDVKLIASASFDSKSVRLWRASDGOYMATFRGHVQAVYTVVAWSADSRLLVSGSKDSTL	433	

Db 367 VFSPDTRYIASAFDKSVKLMWCGRTGKYLASLRGHVGPVYQVAWSADSRLLVSGSADSTL 426

QY 434 KVMVSQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKWL 478

Db 427 KVFELTKSLYYDLPGHGDEVFTVDWSPEGTKVVSQKDKVLKLW 471

RESULT 2

S19487

hypothetical protein YCR072c - yeast (Saccharomyces cerevisiae)

C;Species: Saccharomyces cerevisiae

C;Date: 31-Mar-1992 #sequence\_revision 04-Dec-1992 #text\_change 19-Apr-2002

C;Accession: S19487; S26657

R;Ballesta, J.P.G.; Franco, L.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E. submitted to the Protein Sequence Database, March 1992

A;Reference number: S19486

A;Accession: S19487

A;Molecule type: DNA

A;Residues: 1-484, 'I', 486, 'TKL', 490, 492-515 <BAL1>

A;Cross-references: EMBL:X59720; MIPS:YCR072c

A;Note: this sequence has been revised in reference S26657

R;Ballesta, J.P.G.; Franco, L.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E. submitted to the Protein Sequence Database, October 1992

A;Reference number: S26587

A;Accession: S26657

A;Molecule type: DNA

A;Residues: 481-503 <BAL2>

A;Cross-references: EMBL:X59720; MIPS:YCR072c

C;Genetics:

A;Cross-references: SGD:S0000668

A;Map position: 3R

C;Superfamily: unassigned WD repeat proteins; WD repeat homology

F;182-215/Domain: WD repeat homology <WD2>

F;225-265/Domain: WD repeat homology <WD3>

F;274-306/Domain: WD repeat homology <WD4>

F;398-431/Domain: WD repeat homology <WD5>

F;440-473/Domain: WD repeat homology <WD6>

Query Match 42.4%; Score 1092; DB 2; Length 515;

Best Local Similarity 43.3%; Pred. No. 4.3e-77;

Matches 214; Conservative 95; Mismatches 147; Indels 38; Gaps 11;

QY 14 TIQARLVYTGEEAGPPIDLPAGITTOQLGLICNALL-KNEEATPYLFFV-----GEDEI 66

Db 33 SIKFQALDTGDNVGGALRVPGAISEKQLEELLNQLNGTSDDPVPYTFSCATIOKKASDPV 92

QY 67 KKSLEDTLDLAS-----VDTENVIDIYQQAQVFKVRPVTCTSSMPGHAEAVV 115

Db 93 K-----TIDITDNLSSLIKPGYNSTEDQITLLYTPRAVFKVKPVTSSSAIAGHSTIL 147

QY 116 SLNFSP-DGAHLASGSGDTTVRLWDLNTERPHFTCTGKHQWVLCVSWAPDQKRLASGCKA 174

Db 148 CSAFAPHTSSRMVTFGAGDNTARIWDCDTPMPHTLKGHYNWVLCVSWSPDGEVIATGSM 207

QY 175 GSIIWDPETGQKGRPLSGHKKHINCLAWEPYH--RDPECRKLASASGDGDCRIWDVKL 232

Db 208 NTIRLWDPKSGQCLGDALRGHSKWITSLWEPIHLVKGSKPRLASSSKDGTIKIWDTVS 267

QY 233 GQCLMNIAGHTNAVAVRWGGAGLIYTSKKDRTVKMWRA-ADGILCRTFSGHAHWNNIA 291

Db 268 RVCQYTMSGHTNSYCVKWGGQLLYSGSHDRTVVRVWDINSQGRGINILKSHAHWVNHLS 327

QY 292 LSTDYVLRTPGF-HPVKDRSKSHLSLSTELQESALKRYQAVCP---DEVESLVSCSDON 347

Db 328 LSTDYALRIGAFDHTGKKPS-----TPAEAQKALENYEKICKKNGNSEMMVITASDDY 381

QY 348 TLYLWRN-NQNKCVERTGHQNVVNDVKYSPDVKLLIASAFDKSVRLWRASDGQYMATFR 406

Db 382 TMFLWNLKSTKPTARMTGHQVLNVHVAFPDGRYIVSASFDSIKLWDRGRDKFISTFR 441

QY 407 GHVQAVTVVAWSADSRLLIVSGSKDSTLKVWSQTKKLAQELPGHADEVFGVDWAPDGSRV 466

Db 442 GHIASYQVAWSDDRLVSCSKDTTLKVWDVTRKLSVDLPGRKDEVTYVDWSVDGKRV 501

QY 467 ASGGKDKVIKLWAY 480

Db 502 CSGGKDKMVLWTH 515

RESULT 3

T41148

trp-asp repeat containing protein - fission yeast (Schizosaccharomyces pombe)

C;Species: Schizosaccharomyces pombe

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 26-May-2000

C;Accession: T41148

R;Hilbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G. submitted to the EMBL Data Library, October 1998

A;Reference number: Z21973

A;Accession: T41148

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-502 <HIL>

A;Cross-references: EMBL:AL031907; PIDN:CAA21419.1; GSPDB:GN00068; SPDB:SPCC18.05c

A;Experimental source: strain 972h-; cosmid c18

C;Genetics:

A;Gene: SPDB:SPCC18.05c

A;Map position: 3

A;Introns: 68/3; 200/3; 299/3

C;Superfamily: unassigned WD repeat proteins; WD repeat homology

Query Match 40.7%; Score 1047.5; DB 2; Length 502;

Best Local Similarity 45.3%; Pred. No. 1.2e-73;

Matches 208; Conservative 76; Mismatches 160; Indels 15; Gaps 6;

QY 32 LPAGITTOQLGLICNALLKN-EEATPYLFFVGDEIKKSLEDTLDLA-----SVDTENVI 85

Db 49 VPGNSSVRQLEALLNQLLENSDDPPVPYNFALHHEDETIEIQDNLYTSVFHGLMKTEDHL 108

QY 86 DIVYQQAQVFKVRPVTCTSSMPGHAEAVVSLNFSPD-GAHLASGSGDTTVRLWDLNTER 144

Db 109 TLLYTPQAVFRVRAVTRCTASMNHGDGTIIISAQFSPSTSSRLVTGSGDFTARLWDCDTQT 168

QY 145 PHFTCTGHKQWVLCVSWAPDQKRLASGCKAGSIIWDPETGQKGRPLSGHKKHINCLAW 204

Db 169 PIATMKGHTNWVSCVWAPDASIIATGSMDNTRIFWDPKKGSPIGDALRHTKPIMALCW 228

QY 205 EPHYHROPECRK--LASASGDGDCRIWDVKLGQCLMNIAGHTNAVAVRWGGAGLIYTSKK 262

Db 229 QPLHLAPDSGPYLLASGSKDNTVRIWNVKLRTLTLTSLGHTAPITCVRWGGQNWIYSSY 288

QY 263 DRTVKMRAADGILCRTFSGHAHWNNIALSTDYVLRTPGPHVPKDRSKSHLSLSTELQ 322

Db 289 DKTIRIWDADGKCLHLKGAARVNHLSLSTELVLRSGAYDHTDFKPKSF-----SDER 343

QY 323 ESALKRYQAVCPDEVESLVSCSDDNLYLW-RNNQNKCVERTGHQNVVNDVKYSPDVKL 381

Db 344 RKAKERYEACLKQSGERLVASDDLLQLILWDPOKSTKPTIKMHGQKVYNHASFSPDGR 403

QY 382 IASAFDKSVRLWRASDGQYMATFRGHVQAVTVVAWSADSRLLIVSGSKDSTLKVWSYQTK 441

Db 404 IATASFDSVRLWDGKTGKFLATLRGHVAAVYQCAWSTDSRLLVSSSQDTTLKVWDVRSK 463

QY 442 KLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKLWAY 480

Db 464 KMKFDLPGHEDQVFAVDWSPDQQRVASGGADKAVRIWSH 502

RESULT 4

A12155

WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp.

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C;Accession: A12155

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.;

DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AF2071  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1683 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA000019; PID:g17131215; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr2800

Query Match 23.2%; Score 597; DB 2; Length 1683;  
Best Local Similarity 26.2%; Pred. No. 7.6e-38;  
Matches 149; Conservative 91; Mismatches 174; Indels 154; Gaps 9;  
QY 103 CTSSMPGHAEAVVSLNFSFDPGHAHLASGSGDITVRLWDLNLTETPHFTCTGKHQWVLCVSWA 162  
DB 718 CIKTLTGHHEVFSVAFHPDGTETLASAGDKTIKLDIQDGTCLTGTHTDWRVCVAFS 777  
QY 163 PDGKRLASGCKAGSIIWDPETGQKGRPLSGHKKHINCLAWEPYHRDPECKRLASAGD 222  
DB 778 PDGNTLASSAADHTIKLDVSGQKCL-RTLKSHGTGWVRSVAF-----SADGQTLASGSGD 831  
QY 223 GDCRIWDVKLGQCLMNIAGHTNAVTAVRWG-GAGLIYTSSKDRTVKMWRAADGILCRTFS 281  
DB 832 RTIKIWNHYHTGECLKTYIGHTNSVSIAYSPDSKILVSGSGDRTIKLDWCQTHICIKTLH 891  
QY 282 GHAWVNNIALSTD-----YVLRTPGFHPVKDRSKSHLSLSTEELQESAL 326  
DB 892 GHTNEVCVAFSPDGQTLACVSLDQSVRLWNCRTG-----QCL 929  
QY 327 KRYQ-----AVCPDESVLSCSDNTLYLWRNNQKCVERTGHQNVVNDVKYSP 377  
DB 930 KWTGNTDVALPVAFPDR-QILASGSNDKTVKLDWQTKYISSLEGHTDFIYGIAFSP 988  
QY 378 DVKLIASAFDKSVRLWRASDQYMATFRGHVQAVTVVAVSADSRLLIVSGSKDSTLKVWS 437  
DB 989 DSQTLASASTDSSVRLWNIQTGCFQILLEHTDWTAVVAFHPQCKIATGSADCTVKLWN 1048  
QY 438 VQTKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKWLW 478  
DB 1049 ISTGQCLKTLSEHSDKILGMAWSPDGQLLASASADQSVRLW 1089

RESULT 5  
AF2071  
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp.  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C;Accession: AF2071  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AF2071  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1683 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA000019; PID:g17131215; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr2800

Query Match 23.2%; Score 597; DB 2; Length 1683;  
Best Local Similarity 26.2%; Pred. No. 7.6e-38;  
Matches 149; Conservative 91; Mismatches 174; Indels 154; Gaps 9;  
QY 36 ITTQOGLGLICNALKNEEATPYLFFVGEDEIKKSLDTELTLDLASVDTENVIDIVYQPQAVF 95  
DB 1019 LSHQQLAALIASLKAAQVNHVI-----AVPNNKLATVTT-----LQQAALF 1060

QY 96 KVRPYTRCTSSMPGHAEAVVSLNFSFDPGHAHLASGSGDITVRLWDLNLTETPHFTCTGKHQW 155  
DB 1061 EMQERNR----LEGHKDGVISISISRDGQTIASGSLDKTIKLSRDGRL-FRTLNGHEDA 1115  
QY 156 VLVSWAPDGRKRLASGCKAGSIIWDPETGQKGRPLSGHKKHINCLAWEPYHRDPECKR 215  
DB 1116 VYSVFSFDPGQTIASGSGDKTIKLTQTSQDGTLL-KTITGHEQTVNNV-----YFSPDGKN 1169  
QY 216 LASASGDGDCRIWDVKLGQCLMNIAGHTNAVTAVRWGGAG-LIYTSSKDRTVKMWRAADG 274  
DB 1170 LASASSDHSIKLWDTTSGQLLMTLTGHSAGVITVRFSPDGTIAAGSEDKTVKLVHRQDG 1229  
QY 275 ILCRTFSGHAHWNNIALSTD----- 295  
DB 1230 KLLKTLNGHQDWVNSLSFSPDGKTLASASADKTIKLVRIADGKLVKTLKGHNDSVWDVNF 1289  
QY 296 -----YVLRTPGFHPVKDRSKSHLSLSTEELQESALK 327  
DB 1290 SSDGKAIASASRDNTIKLWNRHGIELEFTGCHSGGVAVNFPLPSNIIASASLDNTIRLW 1349  
QY 328 RYQAVCPDEV----- 337  
DB 1350 QRPLISPLEVLACNSGVYAVSFLHDGSIATAGADGNIQLWHSQDGLLKLTPGNKAIYG 1409  
QY 338 -----ESLVSCSDNTLYLWRNNQKCVERTGHQNVVNDVKYSPDVKLIASAFDKS 390  
DB 1410 ISFTPGDLIASANADKTVKIWRVRDGAALKTLIGHNEVKNVFPSPDGKTLASASRDNT 1469  
QY 391 VRLWRASDQYMATFRGHVQAVTVVAVSADSRLLIVSGSKDSTLKVWSVQTKLAQELPGH 450  
DB 1470 VKLWNVSDGKFKKTLKGHTDEVFWVSEFSPDGKIIASASADKTIRLWDSFSGNLIKSLPAH 1529  
QY 451 ADEVFGVDWAPDGSRVASGGKDKVIKWLW 478  
DB 1530 NDLVYSVNFNPDGSMLASTSADKTVKWLW 1557

RESULT 6  
AC2239  
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp.  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C;Accession: AC2239  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AC2239  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1526 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA000019; PID:g17132599; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr3466

Query Match 22.9%; Score 590; DB 2; Length 1526;  
Best Local Similarity 34.1%; Pred. No. 2.3e-37;  
Matches 132; Conservative 76; Mismatches 153; Indels 26; Gaps 8;  
QY 103 CTSSMPGHAEAVVSLNFSFDPGHAHLASGSGDITVRLWDLNLTETPHFTCTGKHQWVLCVSWA 162  
DB 1066 CLYTLQGHGHTSCVRSVVFSPDGAMLASGGDDQIVRLWDISSGNCLYTLOGYTSWVRFVFS 1125  
QY 163 PDGKRLASGCKAGSIIWDPETGQKGRPLSGHKKHINCLAWEPYHRDPECKRLASAGD 222  
DB 1126 PNGVTLANGSSDQIVRLWD-ISSKKCLYTLOGHTVWVNAVF-----SPDGATLASGSGD 1179  
QY 223 GDCRIWDVKLGQCLMNIAGHTNAVTAVRWGGAG-LIYTSSKDRTVKMWRAADGILCRTFS 281



```

Db      1180 QTVRLWDISSKCLYLILQGHTSWNVSVFNPDPGSTLASGSSDQTVRLWEINSSKCLCTFQ 1239
QY      282 GHAWVNUNIALSTD-YVLRCTGPFHPVKDRSKSHLSLSTEEQLQBSALKRYQ-----A 331
Db      1240 GHTSWNVSVFNPDPGSMILASGS---SDKTVRLWDISSK---CLHTFQGHNTWNVSA 1291
QY      332 VCPDEVESLVSCSDNTLYLWRNNQKCVERTMTHQNVNVNDVKYSPDVKLIASASTDKSV 391
Db      1292 FNPDG-SMLASGSGDQTVRLWEISSKCLHTFQGHTSWVSVTFSPDGTMLASGSDQTV 1350
QY      392 RLWRASGQYMATPRGHVQAYYTVAWSADSRLLVSGSKDSTLKWVSVQTKKLAQELPGHA 451
Db      1351 RLWSSISGECLYTFLLGHTNWWGSVIFSPDGAILASGSDQTVRLWSSISGKCLYTLQGHN 1410
QY      452 DEVFGVDWAPDGSRVASGGKDKVIKWL 478
Db      1411 NWGVSIVFSPDGTLLASGSDQTVRLW 1437

RESULT 7
Tl8521
beta transducin-like protein - Podospora anserina
C;Species: Podospora anserina
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: Tl8521
R;Saupe, S.; Furcq, B.; Begueret, J.
Gene 162, 135-139, 1995
A;Title: A gene responsible for vegetative incompatibility in the fungus Podospora anserina
A;Reference number: 218944; MUID:96009891; PMID:7557402
A;Accession: Tl8521
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1356 <SAU>
A;Cross-references: EMBL:L28125; NID:g607002; PID:g607003; PIDN:AAA85775.1
C;Genetics:
A;Gene: het-el
A;Introns: 761/3

```

QY	452	DEVFGVDNAPDGRSVASGGKDKVIKLW	478
		: :	
Dd	1178	GWVQSVAFPDQGQRVASGSSDKTIKW	1204
RESULT 8			
AEI1866			
WD-40 repeat protein [Imported] - Nostoc sp. (strain PCC 7120)			
C:Species: Nostoc sp.			
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120			
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002			
C:Accession: AEI1866			
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Tai Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tai			
DNA Res. 8, 205-213, 2001			
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacter			
A:Reference number: AB1807; MUID:21595285; PMID:11759840			
A:Accession: AEI1866			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1708 <KUR>			
A:Cross-references: GB:BA000019; PIDN:BAE72436.1; PID:g17129823; GSPDB:GN00179			
A:Experimental source: strain PCC 7120			
C:Genetics:			
A:Gene: all0478			
Query Match                  20.9%;    Score 538.5;    DB 2;    Length 1708;			
Best Local Similarity       32.3%;    Pred. No. 2.8e-33;			
Matches 147; Conservative 77; Mismatches 168; Indels 63; Gaps 15;			
QY	42	GLICNALLNKEEATPYLFFVGEDEIKKSLEDTLDLASVDTENVIDIVYQPAVFVKRPVT	101
		:     :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1137	GSLINTLSKHTNVVNSVNF-----SPDALLIASAQDKTVKLVNR-----VG	1178
QY	102	RCTSSMPGHAAEVVSLNFSPDGAHLASGSGDTTVRLND-----LNTETPHETCTGHKKQWV	156
		: : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1179	QLVTTLQGHGDVVNNASFPSDGLIASGSDTKTVKLSREGKLLN-----TSLGHNDAY	1232
QY	157	LCVSWAPDGKRRLASGCAGSIIIWDPETGQQKGRLP---SGHKKHINCLAWEPYHRDPEC	213
		:       :       :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1233	LGIATPDGOTLASVGADKNIKWN-----RDGILLKTWQHDDAILGVAV-----SPKG	1282
QY	214	RKLASAGDGCRIWDVKVLGOCLMNIAGTNNAVTVRWGGAG-LIYTSSKDRTVKMWRAA	272
		:       : : :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1283	EIATASFDTIKLWN-RQNLLKLTLSGHTAGVTAFTFSPNGETIGSASIDATLKLV-SP	1340
QY	273	DGILCRTEFSGHAHVNNIALSTDYVLRTPGFHPVKDRSKSHLSLSTEELOESALKRYQAV	332
		:                   :   :   :   :   :   :   :   :   :   :   :	
Dd	1341	QGILLGLTKGHNWSVNSVSFSD-----GRIFASGRDK--TJTLWRWDEVLLRNPRGD	1392
QY	333	CPDEVESLWSCDDNTLYLWRNNQ-----NKVERMTGHQNVVNDVKYSPDVKLIIAS	384
		:   :         :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1393	GNDWVTSISFSSDGETLAASRDQTVKILSRHGKULLNTFKHGTGSINGVAVSPNQMIAS	1452
QY	385	ASFDKSVRLWRASDQYMATERGHVQAIVTVANSADSRLLIVSGSKDSTLKVMSVQTKKLA	444
		:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1453	ASKDOTVKLWH-QDGKILHTLQHQDQDLAVAWSSDSQVIASAGKDKIVKIWS-QGGQLL	1510
QY	445	QELPGHADEVFGVDWAPDGRSVASGGKDKVIKLWA	479
		: :   :   :   :   :   :   :   :   :   :   :   :   :	
Dd	1511	HTLOGTDAVNWSVSFDGKLLASVSDTTTVKLLWS	1545



DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE1810  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1227 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA877553.1; PID:g17135007; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr0029

	Query Match	20.7%;	Score 533;	DB 2;	Length 1227;
	Best Local Similarity	32.8%;	Pred. No. 4.7e-33;		
	Matches 131;	Conservative 66;	Mismatches 157;	Indels 46;	Gaps 11;
QY	103	CTSSMPGHAEEAVVYSLNESPDGAHLASGSGDITVRLWDLNTEPHTCTCTGHKQWLCVSWA	162		
Db	678	CLKLTKSNTKNKYVSAFSPDGRILASASODQTIKLWDLATGNCQQTLLIGHDDWVWSVTF	737		
QY	163	P--DGK--RLASGCKAGSIIIWDPETGOOKRPLSKHKKHINCLAWEPYHRDPECKRLAS	218		
Db	738	PVTDDRPLLLASSADQHIKLWDVATGKCL-KTLKGHTREVHSVSF-----SPDGQTLAS	791		
QY	219	ASGDGDCRIWDVKLGQCLMNTAGHTNAVTAVRWGGAGLIYTS-SKDRTVKWWRADGILC	277		
Db	792	SGEDSTVRLWDVKTGCQWQIFEGHSKKYVSFRFSPDQTLASCGEDRSIKLWDIQGECV	851		
QY	278	RTFSGHAHWNNIALSTDYVLRTPGFHPVKDRSKSHLSLSTEEQLQ-----SALK	327		
Db	852	NTLGHSSQVWAIAFSPD-----GRTLISCSDDQARLWDVITGNSLNILR	897		
QY	328	RYQ-----AVCPDEVESLVCSDDNLTLYLRNNQNKCVERTGHQNVVNDVKYSPDVKL	381		
Db	898	GYTRDVIYSAFSPDS-QILASGRDDYITGLWNLKTGEC-HPLRGHQGRIRSAFHPDGKI	955		
QY	382	IASAFDKSVRLWRASD---CQYMATFRGHVQAVYTVAWSADSRILVSGSKDSTLKVWSV	438		
Db	956	LASSGADNTIKLWDISDTHNSKIYIRTLTGHTNNWVTVVFSPDKHTLASSSEDTIRLWDK	1015		
QY	439	QTKKLAQELPGHADEVFGVDWAPDGRSVASGCKDKVTKLW	478		
Db	1016	DTGDCLOKLKGHSHWVTVVAFSPDGRILASGSADSEIKIW	1055		

```

RESULT 10
AD1842
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD1842
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1711 <CUR>
A:Cross-references: GB:BA000019; PIDN:BA077808.1; PID:g17135262; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all0284

```

```

Query Match      20.7%; Score 532.5; DB 2; Length 1711;
Best Local Similarity 28.2%; Pred. No. 8.2e-33;
Matches 156; Conservative 98; Mismatches 183; Indels 117; Gaps 19;

QY      16 QARLVYTGEEAGPPID- - - -LPAGITQQLGLICNALLKNEEATPYLFFVGEDEIKKSL 70
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1057 QLEALMTSVQAGKEQVNIISGIPDTRQTATILQQAQVYNTQERNRLLLNHWVTSVSYSP 1116

```

Qy	71	EDTLDLASVDPTENVIDIVYQOQAVFKVRPVRTCTSSMPGHAAEVVSLNFSPPDCAHLASGS	130
Db	1117	DGEV-IASGSVDNTIHL-----WRRDGKLLTTLTGHDNGVNSVSFSPDGEILASAS	1166
Qy	131	GDTVRLWDLNTPHFTCTGCHKQWVLCVSWAPDGKRLASGCKAGSIIWDPETGQQKGR	190
Db	1167	ADSTIKLWQRNGQL-ITTLKGHDQGVKSVSFSPNGEIIASGSDHTINLW-----SRAGK	1220
Qy	191	---PLSGHKKHINCLAWEPYHRDPECKRLASASGDGDCRIWDVKLGQCLMNTAGHTNAV	247
Db	1221	LLLSLNGHSQGVNSIKF-----SPEGDTIASASDDGTIRLWSLD-GRPLITIPSHTKQVL	1274
Qy	248	AVR-----WGGAG-----LIYTSSKDR	265
Db	1275	AVTFSPDQGTIVSAGADNTVKLSRNGRTLLTLEGHNEAVQVTFSPDGRLLATASADKT	1334
Qy	266	VKMWRAADGILCRTFSGHAHVNNIALSTD-----YVLRGTGP--FHPVK	307
Db	1335	ITLW-SRDGNILGTFAGHNHEVNSLSFSPDGNILASGDDNTVRLWTVNRTPLKTFYGHK	1393
Qy	308	-----DRSKHLSLTE-----ELOESALKRYQAVCPD-----EVESLVSCS	344
Db	1394	GSVSYVRFNSNDGKKITSLSTDSTMTKWSLDGKLLQLTSSPLPDTVSISFTPDNKIVALAS	1453
Qy	345	DDNTLYLWRNNQKCVERTGHQNVNDVKYSPDVKLIASAFDKSVRLWRASDGOYMAT	404
Db	1454	PDHTIHLV-NROGGLLRSLPGHNHWITLSFSFNKQILASGSADKTIKLW-SVNGRLKLT	1511
Qy	405	FRGHVQAVYTVAMSADSRLLIVSGSKDSTLKVWSVQTKKLAQELPGHADEVFGVDWAPDGS	464
Db	1512	LLGHNGWVTDIKFSADGKNIVSASADKTIKIWSLD-GRLIRTLQCHSASVWVNSLSPDQG	1570
Qy	465	RVASGGKDKVKIKLW	478
Db	1571	TLASTSQDETIKLW	1584

RESULT 11  
AC1842  
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AC1842  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.;  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.;  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacter  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC1842  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1747 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA077807.1; PID:gl7135261; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: al10283

	Query Match	20.7%	Score 532;	DB 2;	Length 1747;
	Best Local Similarity	27.0%;	Pred. No. 9.2e-33:		
	Matches 132; Conservative	89;	Mismatches 167;	Indels 100;	Gaps 12;
QY	73	TLDLASVDTENV	DIVYQQAVFKVRPVRTCTSSMPGHAAEVVSLNFSFDPDGAHLASGSGD	132	
		I : : : I I :	I : : : I I I : : I I I I : I I I :		
Db	1201	TIKLYALDTSCLI-----VNLLQCKIQLIKTFEGHTDIVDVVFSPDKSTIVSSSLD	1252		
QY	133	TVRLWDLNTETPHFTCTGHKKQWVLCSWAPDGKRKLASGCKAGSIIIWDPETGQKGRLP	192		
		: I I I I : : I : I I I I I I I I I I : : I : I			
Db	1253	KTIKLWRIDGSIIIN-TWNARHNGWVNSISFSPDGKMIASGGEDNLVKLMQAATNGHLI-KTL	1310		
QY	193	SGHKHHINCIAWEYPHRDPECRKLASASGDGDCRTIWDVKLGQCLMNIAGHTNAVTAVRWG	252		
		: I I I I : : : I : I I I I I I I : : I : I I I I I I I : : I : :			

Db 1311 TGHKERITSVKF-----SPDGKILASASGDKTIKFWNTD-GKFLKTTIAAHNQVNSINFs 1364

Qy 253 G-AGLIYTSKDRTVKMWRAADGILCRTLTFSGHAHWNNI-----ALSTDYVLR TG 301

Db 1365 SDSKTLVSAGADSTMKVWK-IDGTLIKTISGRGEQIRDVTFSPDNKVKLIASASSDKIVRIR 1423

Qy 302 PPHPVKDRSKSHLSLS-----TEELQESALKRYQ-----A 331

Db 1424 QLNYYQKSQKSNVNSVSFNPDKTFASAGWDGNITIWORETLAHSSLSSTIQKNQNIITVS 1483

Qy 332 VCPDEVESLYSCDDNTLYLWRNNQNCVERMTGHQNVVNDVKYSPDVVKLIASASFDKSV 391

Db 1484 YSPDG-KTIATASADNTIKLWDSQTQOLIKTLTGCHKDRITITLSFHPDNQTIASGSADRTI 1542

Qy 392 RLWRASDGQYMATFRGHVQAVYTVAWSADSRLLVSGSKDSTLKV----- 435

Db 1543 KIWRVNDGQLLRTLITGHNDVTSVNFSPDGQFLASGSTDNVTKIWQTDGRLIKNI THGL 1602

Qy 436 -----WSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGG 470

Db 1603 ATASVKFSPDSHTLASASWDNTIKLWQVTDGKLI NNLNNGHIDGVTLSFSFSPDGEILASGS 1662

Qy 471 KDKVIKWLW 478

Db 1663 ADNTIKWLW 1670

RESULT 12

AGI1889

WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AGI1889

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AGI1889

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-934 <KUR>

A:Cross-references: GB:BA000019; PIDN:BAB72622.1; PID:g17130010; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all0664

Query Match 19.3%; Score 497; DB 2; Length 934;

Best Local Similarity 29.7%; Pred. No. 2.1e-30;

Matches 126; Conservative 88; Mismatches 140; Indels 70; Gaps 15;

Qy 106 SMPCHAEAVVSLNFSPDGAHLASGSGDTTVRLWDLNTETPHFTCTG HKQWVLCVSWAPDG 165

Db 445 TLEGHKDKVNSITFSPDGQLIATVGDNTMKLWLDGKELR-TFRGHQDMIWVSFSPDG 503

Qy 166 KRLASGCKAGSIIIDPETGQOKGRPLSGHKHKHINCLAWEPYHRDPECKRLASASGDGDC 225

Db 504 KQIATASGDRTVKLWSLD-GKEL-QTLRGHQGVNSVTF-----SPDGKLIATASGDRTV 556

Qy 226 RIWDVKLGQCLMNIAGHTNAVTAVRWGGAG----- 255

Db 557 KLWNSK-GQELETLYGHTDAVNSVAFSPDGTSIATAGNDKTAKIWLKNSPNSIIVRGHED 615

Qy 256 -----LIYTSKDRTVKMWRAADGIL--CRTFSGHAHWNNIALSTD--YVLR 299

Db 616 EVFDLVFSPNCKYIATASWDKTAKLWSIVGDKLOELRTFNGHGRVKNKLSFSFSPDGKIAT 675

Qy 300 TGPHFPVKDRSKSHLSLSTEELQESALKRYQAV----CPDEVESLYSCSDDNNTLYLWRN 354

Db 676 TS-----WDKTAKLWNLD-GTLQKTLTGHKDTVMSVNFSPDG-QLIATASBEDKTVKLW-N 727

Qy 355 NONKCVERMTGHQNVVNDVKYSPDVVKLIASASFDKSVRLWRASDGQYMATFRGHVQAVYT 414

Db 728 RDGELLKTLPRQSSVNSAVFSPDGKLIATATAGWDKTVKIW-SIDGRLQKTLTGHTSGINS 786

Qy 415 VAWSADSRLLVSGSKDSTIKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDKV 474

Db 787 VTFSPDGKLIASASWDNTVKIWNLDGKEL-RTLGRHKNNVHVNTFSPDGKLIATASGDNT 845

Qy 475 IKLW 478

Db 846 VKIW 849

RESULT 13

AI2493

WD-repeat protein [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120alpha

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AI2493

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; I Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Ta DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacter

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AI2493

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1189 <KUR>

A:Cross-references: GB:BA000020; PIDN:BAB78213.1; PID:g17135667; GSPDB:GN00180

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr7129

A:Genome: plasmid

Query Match 19.3%; Score 496.5; DB 2; Length 1189;

Best Local Similarity 29.3%; Pred. No. 3.2e-30;

Matches 137; Conservative 78; Mismatches 170; Indels 83; Gaps 11;

Qy 72 DTLDIASVDTFENVIDI-----VYQPQ-----AVFKYRPTV----- 101

Db 573 DGCQLATSDTSGVINIWDVNVNGKQLFNCQEHNSWIWDVAFSSVAPVLASCGQDHTIKLMN 632

Qy 102 ----RCTSSMPCHAEAVVSLNFSPDGAHLASGSGDTTVRLWDLNTETPHFTCTG HKQWVL 157

Db 633 TTTGECFNTLHGHTSI VTSVAFSPGKLLASSSYDSHVSKVWDLDTGECLOITFLGHDACVW 692

Qy 158 CVSWAPDGRKRLASGCKAGSIIIDPETGQOKGRPLSGHKHKHINCLAWEPYHRDPECKRKL A 217

Db 693 SVVFHPVGQILATAGEDNTIKLWELQSGCCL-KTLQGHQHWVKTI AF-----NSGGRILA 746

Qy 218 SASGDGDCRIWDVKLGQCLMNIAGHTNAVTAVRWGGA-GLIYTSKDRTVKMWRAADGIL 276

Db 747 SGSPDQNVKLWDIHTGKCVMTLQGHYGVVTSVAFNPKDNLLLSGSDYDQSVKVD RKTGRC 806

Qy 277 CRTFSGHAHWNNIALSTDYVLRTPGFHPVKDRSKSHLSLS-----TE 319

Db 807 LDTLKKHTNRIWSVA-----FHP-----QGHLFVSGGDDHAAKIWELGTGQC IK 850

Qy 320 ELQESALKRYQAVCPDEVESLYSCSDDNNTLYLW-----RNNQNKCV ER-MTGHQNVV 370

Db 851 TFOGHSNATYTTIAHNWEHSHLSLASGHEDQTIKLDLNLHSPHKSNVNTHPFRILQCHSNRV 910

Qy 371 NDVKYSPDVVKLIASASFDKSVRLWRASDGQYMATFRGHVQAVYTVAWSADSRLLVSGSKD 430

Db 911 FSVVFSSTGQLLASGSADRTIKLWSPHTGQC LHTLHGHSWVWALAFSLDDKLLASGSYD 970

Qy 431 STLKVWSVQTKKLAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKWLW 478

Db 971 HTVKIWDVSSGGCQLQTLQGHPGSVLAVAFSCDGKTLFSSGYEKLVKQW 1018

RESULT 14

AB2410





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:52:40 ; Search time 35 Seconds  
(without alignments)  
568.818 Million cell updates/sec

Title: US-09-830-980-1  
Perfect score: 2574  
Sequence: 1 MQETDTEQATPHTIQLRV.....PDGSRVASGGKDKVILWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues 112892

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1092	42.4	515	1	YCW2_YEAST	P25382 saccharomyc
2	599	23.3	1258	1	YS00_ANASP	Q8ytc2 anabaena sp
3	597	23.2	1683	1	YL24_ANASP	Q8yv57 anabaena sp
4	590	22.9	1526	1	YY46_ANASP	Q8yrl1 anabaena sp
5	581	22.6	1356	1	HET1_PODAN	Q00808 podospora a
6	433	16.8	1693	1	YL63_SYNY3	Q55563 synecocyst
7	419	16.3	334	1	WDR5_HUMAN	Q9ugp9 homo sapien
8	415	16.1	361	1	WDS_DROME	Q9v378 drosophila
9	409.5	15.9	742	1	PKWA_THECU	P49695 thermomonos
10	372	14.5	376	1	YKY4_CAEEL	Q17963 caenorhabdi
11	371	14.4	577	1	TBL1_HUMAN	O60907 homo sapien
12	361.5	14.0	501	1	YH92_CAEEL	Q23256 caenorhabdi
13	359	13.9	614	1	TU11_SCHPO	Q09715 schizosacch
14	354	13.8	409	1	LIS1_BOVIN	P43033 bos taurus
15	354	13.8	409	1	LIS1_HUMAN	P43034 homo sapien
16	354	13.8	409	1	LIS1_MOUSE	P43035 mus musculus
17	347.5	13.5	395	1	YZLL_CAEEL	Q93847 caenorhabdi
18	346	13.4	678	1	SCOB_EMBNI	Q00659 emericeella
19	320	12.4	423	1	WDR5_HUMAN	Q9gz17 homo sapien
20	319.5	12.4	542	1	FW1B_HUMAN	Q9ukd1 homo sapien
21	319	12.4	312	1	GBLP_LETMA	Q25306 leishmania
22	318.5	12.4	586	1	TU12_SCHPO	Q9uug8 schizosacch
23	316	12.3	423	1	WDR5_MOUSE	Q9jj44 mus musculus
24	315	12.2	312	1	GBLP_LETCH	Q27434 leishmania
25	314.5	12.2	518	1	TRCB_XENLA	Q91854 xenopus lae
26	313.5	12.2	605	1	FW1A_HUMAN	Q9y297 homo sapien
27	312.5	12.1	605	1	POF1_SCHPO	P87053 schizosacch
28	312.5	12.1	665	1	LI23_CAEEL	Q09990 caenorhabdi
29	312.5	12.1	640	1	MT30_YEAST	P39014 saccharomyc
30	312.5	12.1	608	1	WDR1_DROME	Q9vu68 drosophila
31	312.5	12.1	608	1	WDR1_CANAL	P56093 candida alb
32	312.5	12.1	682	1	TUP1_KLULA	P56094 kluyveromyc
33	312.5	12.1	1248	1	APAF_HUMAN	O14727 homo sapien

34	303	11.8	1249	1	APAF_RAT	Q9epv5 rattus norv
35	301	11.7	606	1	PF20_CHLRE	P93107 chlamydomon
36	300	11.7	604	1	RCO1_NEUCR	P78706 neurospora
37	300	11.7	1249	1	APAF_MOUSE	O88879 mus musculu
38	297	11.5	579	1	SE10_CAEEL	Q93794 caenorhabdi
39	296	11.5	340	1	GBB1_CAEEL	P17343 caenorhabdi
40	296	11.5	609	1	WDR1_CHICK	O93277 gallus gall
41	294.5	11.4	713	1	TUP1_YEAST	P16649 saccharomyc
42	292.5	11.4	404	1	WSB2_HUMAN	Q9nys7 homo sapien
43	291	11.3	704	1	T2D4_DROME	P49846 drosophila
44	290.5	11.3	341	1	GBB_LYMST	Q08706 lymnaea sta
45	290	11.3	650	1	SCO2_NEUCR	Q01277 neurospora

ALIGNMENTS

RESULT 1  
YCW2\_YEAST  
ID YCW2\_YEAST STANDARD; PRT; 515 AA.  
AC P25382;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 57.0 kDa Trp-Asp repeats containing protein in CPR4-SSK22  
DE intergenic region.  
GN YCR072C OR YCR72C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ballista J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,  
RA Sanz E.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RA Jimenez A.;  
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X59720; CAA42270.1; -  
DR PIR; S19487; S19487.  
DR SGD; S0000668; YCR072C.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 8.  
DR PRINTS; PR00320; GPROTEINBRPT.  
DR ProDom; PD000018; WD40; 6.  
DR SMART; SM00320; WD40; 8.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 5.  
DR PROSITE; PS50082; WD\_REPEATS\_2; 7.  
DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
KW Hypothetical protein; Repeat; WD repeat.  
FT REPEAT 141 172 WD 1.  
FT REPEAT 184 214 WD 2.  
FT REPEAT 227 264 WD 3.  
FT REPEAT 276 305 WD 4.  
FT REPEAT 318 387 WD 5.  
FT REPEAT 400 430 WD 6.  
FT REPEAT 442 472 WD 7.  
FT REPEAT 484 514 WD 8.  
SQ SEQUENCE 515 AA; 57040 MW; DFE603B79BFB530A CRC64;  
Query Match 42.4%; Score 1092; DB 1; Length 515;



[illegible]

```

RESULT 2
YS00_ANASP
ID YS00_ANASP STANDARD; PRT; 1258 AA.
AC Q8YTC2;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein alr2800.
GN ALR2800.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; Pubmed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: CONTAINS 15 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC

```

DR	EML1; AP003590; BAB74499.1; -.	DR
DR	InterPro; IPR0000767; Disease_resist.	DR
DR	InterPro; IPR001680; WD40.	DR
DR	Pfam; PF00400; WD40; 14.	DR
DR	PRINTS; PRO0364; DISEASERSIST.	DR
DR	PRINTS; PRO0320; GPROTEINBRPT.	DR
DR	ProDom; PD000018; WD40; 1.	DR
DR	SMART; SM00320; WD40; 14.	DR
DR	PROSITE; PS00678; WD_REPEATS_1; 9.	DR
DR	PROSITE; PS00882; WD_REPEATS_2; 14.	DR
DR	PROSITE; PS0294; WD_REPEATS_REGION; 1.	DR
KW	Hypothetical protein; Repeat; WD repeat; Complete proteome.	
FT	Hypothetical protein; Repeat; WD repeat; Complete proteome.	
FT	REPEAT 55 93	
FT	REPEAT 640 679	WD 1.
FT	REPEAT 682 721	WD 2.
FT	REPEAT 724 763	WD 3.
FT	REPEAT 766 807	WD 4.
FT	REPEAT 809 849	WD 5.
FT	REPEAT 850 889	WD 6.
FT	REPEAT 892 931	WD 7.
FT	REPEAT 934 975	WD 8.
FT	REPEAT 976 1017	WD 9.
FT	REPEAT 1019 1059	WD 10.
FT	REPEAT 1060 1101	WD 11.
FT	REPEAT 1103 1143	WD 12.
FT	REPEAT 1144 1183	WD 13.
FT	REPEAT 1186 1227	WD 14.
FT	REPEAT 1258 AA; 139513 MW; 45DF03B91170C451 CRC64;	WD 15.
SO	SEQUENCE	

Query Match	23.3%;	Score 599;	DB 1;	Length 1258;
Best Local Similarity	32.7%;	Pred. No. 7.5e-39;		
Matches 131; Conservative	66;	Mismatches 150;	Indels 54;	Gaps 7;

Qy	103	CTSSMPGHAEEAVVYSLNFSPDCAHLASGSGDTTVRLWDLNTETPHFTCTGKHQWLCVSWA	162
Db	718	CIKTLTGHEHEVFSVAFHPDGETILASASGDKTIKLWDIQDGTCLQTLTGHTDWRVCAFS	777
Qy	163	PDGKRLASGCKAGSIIIWDPETGQKGRPLSGHKHKNCLAWEPYHRDPCECRKLASASGD	222
Db	778	PDGNTLASSAADHTIKLWVSQKGKCL-RTLKSHGTGWRVSAF-----SADGQTLASGSGD	831
Qy	223	GDGRIVDKLGGCLMNIAGHTNAVTAVRWG-GAGLIYTSSKDRTVKMWRAADGILCRTFS	281
Db	832	RTIKIWNHYHTGECLKTYIIGHTNSVYSIAYSKILVSGSGDRTIKLWDCQTHICIKTLH	891
Qy	282	GHAHVWNIALSTD-----YVLRGTGPHFVKDRSKSHLSLSTEEQLQESAL	326
Db	892	GHTNEVCSVAFSPDGQTLACVSLDQSVRLWNCRTG-----QCL	929
Qy	327	KRYQ-----AVCPDEVESLVSCSDONTLYLWRNNQNKCVERTGHQNVVNDVKYSP	377
Db	930	KANYGNTDWALPVAFSPDR-QILASGSDNKTVKLWDQGTGKYISSLEGHTDFIYCIAFSP	988
Qy	378	DVKLIASAFDKSVRLWRASDGOYMATFRGHVQAVYTVAWSADSRILVSGSKDSTLKVWS	437
Db	989	DSQTLASASTDSSVRLWNIISGQCQFQILLEHTDWVYAVVEFHPQGGKIATGSADCTVKLWN	1048
Qy	438	VQTKKLAQELPGHADEVFGVDWAPDGSRVASGGDKVKIKLW	478
Db	1049	ISTGCKLTLSHSDKILGMAWSPDGOILLASASADQSVRLW	1089

RESULT 3		
YL24_ANASP	STANDARD;	PRT; 1683 AA.
ID YL24_ANASP		
AC Q8YV57;		
DT 15-JUN-2002	(Rel. 41, Created)	
DT 15-JUN-2002	(Rel. 41, Last sequence update)	
DT 15-JUN-2002	(Rel. 41, Last annotation update)	
DE Hypothetical WD-repeat protein all2124.		
GN ALI2124.		
OS Anabaena sp. (strain PCC 7120).		
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc		

```
OX NCBI_TaxID=103690;
RN [1]
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: CONTAINS 13 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003588; BAB73823.1; -
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 14.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 14.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS0082; WD_REPEATS_2; 14.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 1070 1109
FT REPEAT 1111 1150
FT REPEAT 1153 1192
FT REPEAT 1195 1236
FT REPEAT 1280 1318
FT REPEAT 1320 1359
FT REPEAT 1361 1400
FT REPEAT 1402 1441
FT REPEAT 1444 1483
FT REPEAT 1486 1525
FT REPEAT 1528 1567
FT REPEAT 1570 1609
FT REPEAT 1611 1650
SQ SEQUENCE 1683 AA; 185459 MW; F7CB361FF54F7137 CRC64;

Query Match      23.2%; Score 597; DB 1; Length 1683;
Best Local Similarity 26.2%; Pred. No. 1.6e-38;
Matches 149; Conservative 91; Mismatches 174; Indels 154; Gaps 9;

QY 36 ITTQQLGLICNALLKNEATPYLFFVGEDEIKSLEDYLDASVDTEVIDIVYQQAVF 95
DB 1019 LSHQQLAALIASLKAAQOVNHI-----AVPNNKLATVTT-----LQQALF 1060

QY 96 KYRPVTRCTSSMPGHAENVSLNFSFDPDGAHLASGSGDTTVRLWDLNTEPHFTCTGHKQW 155
DB 1061 EMQERNR-----LEGHKDGVISISISRDGQTIASGSLDKTKLWSRDGRL-FRTLNGHEDA 1115

QY 156 VLCVSWAPDGKRLASGCKAGSIIIWDPETGQOKGRPLSGHKKHINCLAWEPYHRDPCRK 215
DB 1116 VYVSFSPDQGTIASGSGDKTIKLTQTSNLTLL-KTITGHEQTVNNV-----YFSPDGKN 1169

QY 216 LASASGDCGRIDVYKLGQCLMNIAGHTNAVAVRWGGAG-LIYTSSKDRYTKMWRADG 274
DB 1170 LASASSDHSIKLWDTTSGQLLMTLTGHSAGVITVRFSPDGTIAAGSEDKTVKLWHRQDG 1229

QY 275 ILCRTFSGHAHWNNIALSTD----- 295
DB 1230 KLLKTLNGHQDWVNSLSFSPDGKTLASASADTKIKLWRIADGKLVKTLKGHNSVWDVNF 1289

QY 296 -----YVLRGTFPHVPYKDRSKSHLSLSTEELQESALK 327
DB 1290 SSDGKAIASASRDNTIKLWNRHGIELETFTGHSAGVAVNPLPDSNIIASASLDNTIRLW 1349
```

```
QY 328 RYQAVCPDEV----- 337
DB 1350 QRPLISPLEVLAGNSGVYAVSFLHDGSIATAGADGNIQLMHSQDGLSLKTLPGNKAIYG 1409

QY 338 -----ESLVSCSDDNTLYLWRNNQKCVERTGHQNVVNDVKYSPDVKLIASASFDKS 390
DB 1410 ISFTPGDLIASANADKTVKIWRVRDGGKALKTLIGHDNEVKNVNFSPDGKTLASASRDNT 1469

QY 391 VRLWRASDGOYMATFRGHVQAVYTVAWSADSRLLIVSGSKDSTLKVSWVQTKKLAQELPGH 450
DB 1470 VKLWNVSDGKPKKTLKGHTDEVFWVSFSPDGKTLIASASADTKIRLWDSFSGNLIKSLPAH 1529

QY 451 ADEVFGVDWAPDGSRVASGGDKVVKLW 478
DB 1530 NDLVYSVNFNPDGSMMLASTSADKTVKILW 1557

RESULT 4
YY46_ANASP
ID YY46_ANASP STANDARD; PRT; 1526 AA.
AC Q8YR11;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical WD-repeat protein alr3466.
GN ALR3466.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
CC -!- SIMILARITY: CONTAINS 16 WD REPEATS (TRP-ASP DOMAINS).
CC -!- SIMILARITY: CONTAINS 1 PENTAPEPTIDE REPEAT DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP003593; BAB75165.1; -
DR InterPro; IPR001646; Speptide_repeat.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00805; Pentapeptide; 1.
DR Pfam; PF00400; WD40; 15.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 8.
DR SMART; SM00320; WD40; 15.
DR PROSITE; PS00678; WD_REPEATS_1; 11.
DR PROSITE; PS0082; WD_REPEATS_2; 15.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat; Complete proteome.
FT REPEAT 334 376
FT DOMAIN 823 862
FT REPEAT 862 901
FT REPEAT 904 945
FT REPEAT 946 985
FT REPEAT 988 1027
FT REPEAT 1030 1069
FT REPEAT 1072 1111
FT REPEAT 1114 1153
```

FT REPEAT 1156 1195 WD 9.  
FT REPEAT 1198 1237 WD 10.  
FT REPEAT 1240 1279 WD 11.  
FT REPEAT 1282 1321 WD 12.  
FT REPEAT 1324 1363 WD 13.  
FT REPEAT 1366 1405 WD 14.  
FT REPEAT 1408 1447 WD 15.  
FT REPEAT 1450 1491 WD 16.  
SQ SEQUENCE 1526 AA; 170376 MW; D5DF27509B4A738 CRC64;

Query Match 22.9%; Score 590; DB 1; Length 1526;  
Best Local Similarity 34.1%; Pred. No. 4.9e-38;  
Matches 132; Conservative 76; Mismatches 153; Indels 26; Gaps 8;

QY 103 CTSSMPGHAFAVSLNFS PDGAHLASGSDTVRLWDLNTPHFTCTGKHQWVLCVSWA 162  
Dd 1066 CLYTLOGHTSCVRSVVFSPDGMALASGGDDQIVRLWDISSGNCLYTLQGYTSWVRFVFS 1125

QY 163 PDGKRLASGCKAGSIITWDPETGQKGRPLSGHKHINCLAWEPYHRDPECRKLASASGD 222  
Dd 1126 PNGVTLANGSSDQIVRLWD-ISSKCLYTLQHTNWNNAVF-----SPDGATLASGSGD 1179

QY 223 GDCRIWDVKGQCLMNIAGHTNAVAVRWGGAG-LIYTSSKDRTVKMWRAADGILCRTEFS 281  
Dd 1180 QTVRLWDISSKCLYTLQGHTSWVNSVFNPDGSTLASGSDQTVRLWEINSSKCLCTQTV 1239

QY 282 GHAWVNNIALSTD-YVLRTPGFPHVPRKDRSKSHLSLSTEELQESALKRYQ-----A 331  
Dd 1240 GHTSWNSVFNPDGSMLASGS-----SDKTVRLWDISSK-----CLHTFQGHNTWVNSVA 1291

QY 332 VCPDEVESLVSCDDNTLYLWRNNQKCVERTGHQNVNDVKYSPDVKLIASAFDKSV 391  
Dd 1292 FNPDG-SMLASGSDQTVRLWEISSKCLHTFQGHTSWVSVTFSPDGTMLASGSDQTV 1350

QY 392 RLWRASDGOYMATFRGHVQAVYTVAMSADSRLLIVSGSKDSTLKVMSVOTKKLAQELPGHA 451  
Dd 1351 RLWSISSGCLYTLFGLHTNWNVGSVIFSPDGAHLASGSDQTVRLWSISSGKCLYTLQGHN 1410

QY 452 DEVEGVDMAPDGSRAVSGGKDKVTKLW 478  
Dd 1411 NWVGSIVFSPDGTLLASGSDQTVRLW 1437

RESULT 5  
HETL\_PODAN  
ID HETL\_PODAN STANDARD; PRT; 1356 AA.  
AC Q00808;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Vegetative incompatibility protein HET-E-1.  
GN HET-E1.  
OS Podospora anserina.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Lasiosphaeriaceae; Podospora.  
OX NCBI\_TaxID=5145;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96009891; PubMed=7557402;  
RA Saube S., Turcq B., Begueret J.;  
RT "A gene responsible for vegetative incompatibility in the fungus  
RT Podospora anserina encodes a protein with a GTP-binding motif and G  
RT beta homologous domain."  
RL Gene 162:135-139(1995).  
CC -!- FUNCTION: RESPONSIBLE FOR VEGETATIVE INCOMPATIBILITY THROUGH  
CC SPECIFIC INTERACTIONS WITH DIFFERENT ALLELES OF THE UNLINKED GENE,  
CC HET-C.  
CC -!- SIMILARITY: CONTAINS 10 WD REPEATS (TRP-ASP DOMAINS).  
CC -!- SIMILARITY: CONTAINS 1 NACHT DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
Dd EMBL; L28125; AA85775.1; -.  
Dd InterPro; IPR001680; WD40.  
Dd Pfam; PF00400; WD40; 10.  
Dd PRINTS; PR00320; GPROTEINBRPT.  
Dd ProDom; PD000018; WD40; 10.  
Dd SMART; SM00320; WD40; 10.  
Dd PROSITE; PS50837; NACHT; 1.  
Dd PROSITE; PS00678; WD\_REPEATS\_1; 10.  
Dd PROSITE; PS50082; WD\_REPEATS\_2; 10.  
Dd PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
KW GTP-binding; Repeat; WD repeat.  
FT DOMAIN 294 629 NACHT.  
FT NP\_BIND 300 307 GTP (POTENTIAL).  
FT REPEAT 839 869 WD 1.  
FT REPEAT 881 911 WD 2.  
FT REPEAT 923 953 WD 3.  
FT REPEAT 965 995 WD 4.  
FT REPEAT 1007 1037 WD 5.  
FT REPEAT 1049 1079 WD 6.  
FT REPEAT 1091 1121 WD 7.  
FT REPEAT 1133 1163 WD 8.  
FT REPEAT 1175 1205 WD 9.  
FT REPEAT 1217 1247 WD 10.  
SQ SEQUENCE 1356 AA; 149765 MW; 965FB319844E0651 CRC64;

Query Match 22.6%; Score 581; DB 1; Length 1356;  
Best Local Similarity 27.2%; Pred. No. 2.1e-37;  
Matches 138; Conservative 81; Mismatches 182; Indels 106; Gaps 7;

QY 13 HTIQARLVYTGEEAGPPIDLPAGITFQQLGLICNALLKNEEATPY--LFFVGEDEIKKS 69  
Dd 763 HTIRGRLLIAIVRDG-----YRFALSVRMIIIEKAPLOAYTSALVFAPTDSEIKK 810

QY 70 LEDTLDLASVDTENVIDIVYQPAVFKVRPVTCTSSMPGHAFAVSLNFS PDGAHLASG 129  
Dd 811 IFKKEEPGWISTISVVEAEW-----NACTQTLEGHGSVLSVAFSPDQQRVASG 859

QY 130 SGDTTVRLWDLNTPHFTCTGKHQWVLCVSWAPDGRKRLASGCKAGSIITWDPETGQK 189  
Dd 860 SDDKTIKIWDTASGTGTQTLLEGHGSVLSVAFSPDQQRVASGSDDKTIKIWDAAAGTCT- 918

QY 190 RPLSGHKHINCLAWEPYHR-----DPE 212  
Dd 919 QTLLEGHGRVQSVAFSPDQQRVASGSDDHHTIKIWDAAAGTCTQTLLEGHGSVLSVAFSPD 978

QY 213 CRKLASASGDGDCRIWDVKGQCLMNIAGHTNAVAVRWGGAG-LIYTSSKDRTVKMWRA 271  
Dd 979 GORVASGSGDKTIKIWDTASGTCTQTLLEGHGSVLSVAFSPDQQRVASGSDDKTIKIWD 1038

QY 272 ADGILCRTEFSGHAHVNNIALSTDYVLRTPGFPHVPRKDRSKSHLSLSTEELQESALKRYQA 331  
Dd 1039 ASGTCTQTLLEGHGWVQSVVFSFD----- 1062

QY 332 VCPDEVESLVSCDDNTLYLWRNNQKCVERTGHQNVNDVKYSPDVKLIASAFDKSV 391  
Dd 1063 -----GORVASGSDDHHTIKIWDAAAGTCTQTLLEGHGSVLSVAFSPDQQRVASGSDGTI 1117

QY 392 RLWRASDGOYMATFRGHVQAVYTVAMSADSRLLIVSGSKDSTLKVMSVOTKKLAQELPGHA 451  
Dd 1118 KIWDAAAGTCTQTLLEGHGSVLSVAFSPDQQRVASGSDGTIKIWDAAAGTCTQTLLEGHG 1177

QY 452 DEVEGVDMAPDGSRAVSGGKDKVTKLW 478  
Dd 1178 GWVQSVAFSPDQQRVASGSDDKTIKIW 1204

RESULT 6  
Y163\_SYNY3















```

Qy 408 HVQAVYTVAWSADSR|--IVSGSKDSTLKVMSVQTKKLAQLPCHADEVFGVDWAPDGS 466
      |   :   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db 423 HKNKCYCLFANMSVPLGKHIIISGEDGRILVWSIQTKQIQAILEGHTTPVLATDSHPTLN 482
      |   :   |   :   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Qy 465 RVASGG--KDKVIKLW 478
      :|||  | |||:|
Db 483 IIASGGLEPDNVIRIW 498
      :|||  | |||:|

RESULT 13
TUI1_SCHPO STANDARD; PRT; 614 AA.
AC Q097L5;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcriptional repressor tup11.
GN TUP11 OR SPAC18B11.10.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckie E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Stevens S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Mostl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=20036803; PubMed=10567571;
RA Mukai Y., Matsuo E., Roth S.Y., Harashima S.;
RT "Conservation of histone binding and transcriptional repressor
functions in a Schizosaccharomyces pombe Tuplp homolog.";
RL Mol. Cell. Biol. 19:8461-8468(1999).
CC CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR.
CC CC -!- SUBUNIT: BINDS TO HISTONES H3 AND H4.
CC CC -!- SIMILARITY: CONTAINS 7 WD REPEATS {TRP-ASP DOMAINS}.
CC CC -!- SIMILARITY: BELONGS TO THE TUP1 FAMILY OF WD-REPEAT PROTEINS.
CC CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
```

```

CC -----
DR EMBL; Z50728; CAA90594.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 5.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Transcription regulation; Repressor; Repeat; WD repeat.
FT REPEAT 311..340 WD 1.
FT REPEAT 358..388 WD 2.
FT REPEAT 400..430 WD 3.
FT REPEAT 441..471 WD 4.
FT REPEAT 482..512 WD 5.
FT REPEAT 536..566 WD 6.
FT REPEAT 578..608 WD 7.
SQ SEQUENCE 614 AA; 67331 MW; B02289556DE2C1B1 CRC64;

Query Match 13.9%; Score 359; DB 1; Length 614;
Best Local Similarity 25.5%; Pred. No. 1.7e-20;
Matches 115; Conservative 65; Mismatches 145; Indels 126; Gaps 15;

QY 53 EAPYLFVFGVEDEIKKSLEDTDLASVDTENVIDIVQP--QAVEKVRVPVTRCTSSMPGH 110
Db 260 QAGPDGEYQGREI-APVSDT-EAARKTTSQSWYVTYNPACRKRVENINLVHTLE-----H 312

QY 111 AEAVVSLNFPDGAHLASGS-----DPTVRLWDLNTETPHFTCTGKHQWVLCVSWA 162
Db 313 PSVVCCKVFSNNGKYLATGCNQANVFDTQTKKFLTHESPD--PSRLYVRTIAFS 369

QY 163 PDGKRLASGCKAGSIIIDPETGQOKGR-PLSGHKKHINCLAWEPYHRDPECKLIASAG 221
Db 370 PDGKYLVTGTEDRQIKLWDLST--QKVRYVFSGHEQDIYSLDFS--HNG---RFIVSGSG 422

QY 222 DGCRIWDVKLGOCILMIAGHTNAVTAVRWG-GAGLIYTSSKDRTVKMWRAADGILCRTF 280
Db 423 DRTARLWDVETQCILKLE-IENGVTATAISPNDQFIAGVSLDQIIRVW-----470

QY 281 SGHAHWVNIALSTDYVLRTPFPHPVKDRSKSHLSLSTELQESALKRYQAVCPDEVESL 340
Db 471 -----470

QY 341 VSCSDNTLYLWRNNQNKCVERTMTHQNVNDVKYSPDVKLIASAFDKSVRLWRAS---397
Db 471 ---SVSGTL-----VERLEGHKESVYSIAFSPDSSILLSGSLDKTIKWELQATR 517

QY 398 -----DGOYMATFRGHVQAVYTVAVSADSRLLIVSGSKDSTLKWWSVQTKKLAQELP 448
Db 518 SVGLSAIKPEGICKATYTGHTDFVLSVAVSPDSRWGLSGSKDRSMQFMDLQTGQSYLTQC 577

QY 449 GHADVEFVGVDWAPDGSRVASGGKDKVIKLWA 479
Db 578 GHRNSVISVCFSPDRGQRFASGSGDLRARIWS 608

RESULT 14
LIS1_BOVIN
ID LIS1_BOVIN STANDARD; PRT; 409 AA.
AC P43033;
DT 01-NOV-1995 (Rel. 32, Created)
DI 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Platelet-activating factor acetylhydrolase IB alpha subunit
DE (EC 3.1.1.47) (PAF acetylhydrolase 45 kDa subunit) (PAF-AH 45 kDa
DE subunit) (PAF-AH alpha) (PAFAH alpha) (Lissencephaly-1 protein) (LIS-
DE 1).
GN PAFAH1B1 OR PAFAHA OR LIS1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.

```









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2003, 06:17:40 ; Search time 59 Seconds  
(without alignments)  
1676.316 Million cell updates/sec

Title: US-09-830-980-1  
Perfect score: 2574  
Sequence: 1 MQETDTEQEATPHTIQARLV.....PDGSRVASGGKDKVIKLWAY 480

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2566	99.7	480	5 O96995	O96995 drosophila
2	2554.5	99.2	481	5 Q9VPR4	Q9vpr4 drosophila
3	2554.5	99.2	488	5 Q8T4A2	Q8t4a2 drosophila
4	1520	59.1	485	4 Q9NVX2	Q9nvx2 homo sapien
5	1520	59.1	487	4 Q9BU54	Q9bu54 homo sapien
6	1513	58.8	476	13 O93531	O93531 xenopus lae
7	1512	58.7	485	11 Q8VEJ4	Q8vej4 mus musculu
8	1335.5	51.9	473	10 Q9FLX9	Q9flx9 arabidopsis
9	1331	51.7	447	10 Q9AYI7	Q9ayi7 oryza sativ
10	1318.5	51.2	473	5 Q9TYV3	Q9tyv3 caenorhabdi
11	1047.5	40.7	502	3 O74855	O74855 schizosacch
12	616	23.9	1356	3 Q8X1P4	Q8xlp4 podospora a
13	603	23.4	1356	3 Q8X1P5	Q8xlp5 podospora a
14	603	23.4	1356	3 Q8X1P3	Q8xlp3 podospora a
15	595	23.1	1376	3 Q8X1P2	Q8xlp2 podospora a
16	538.5	20.9	1708	16 Q8YZI2	Q8yzi2 anabaena sp

17	533	20.7	1227	16 Q8ZOR1	Q8zor1 anabaena sp
18	532.5	20.7	1711	16 Q8Z019	Q8z019 anabaena sp
19	532	20.7	1747	16 Q8Z020	Q8z020 anabaena sp
20	497	19.3	934	16 Q8Y223	Q8y223 anabaena sp
21	496.5	19.3	1189	16 Q8YL09	Q8yl09 anabaena sp
22	486	18.9	1551	16 Q8YMU3	Q8ymu3 anabaena sp
23	476.5	18.5	676	16 Q8YSG6	Q8ysg6 anabaena sp
24	469	18.2	1189	16 Q8YTD1	Q8ytd1 anabaena sp
25	462.5	18.0	1241	2 Q9XBD8	Q9xbd8 amycolatops
26	451	17.5	677	16 Q8YZL9	Q8yzl9 anabaena sp
27	436	16.9	304	16 Q8Z054	Q8z054 anabaena sp
28	425.5	16.5	1049	2 Q9ZEM4	Q9zem4 streptomyce
29	425.5	16.5	1676	16 Q9KXX9	Q9kxx9 streptomyce
30	424.5	16.5	559	16 Q8YSC0	Q8ysc0 anabaena sp
31	412.5	16.0	700	5 Q9XZK1	Q9xzk1 drosophila
32	411	16.0	353	5 Q8T776	Q8t776 branchiosto
33	410.5	15.9	317	10 Q9M2Z2	Q9m2z2 arabidopsis
34	400	15.5	589	16 Q8YN14	Q8yn14 anabaena sp
35	399.5	15.5	414	5 Q8SW59	Q8sw59 encephalito
36	399	15.5	330	4 Q9NUL4	Q9nul4 homo sapien
37	396	15.4	613	10 Q9FN19	Q9fn19 arabidopsis
38	389	15.1	411	5 Q96698	Q96698 drosophila
39	386.5	15.0	1051	17 Q8TMS3	Q8tms3 methanosarc
40	385.5	15.0	328	11 Q9D7H2	Q9d7h2 mus musculu
41	377	14.6	522	4 Q9BQ87	Q9bq87 homo sapien
42	372.5	14.5	514	4 Q9BZK7	Q9bzck7 homo sapien
43	372.5	14.5	514	4 Q9H9A1	Q9h9a1 homo sapien
44	371	14.4	786	16 Q8YNK6	Q8ynk6 anabaena sp
45	370.5	14.4	514	4 Q9H2I1	Q9h2i1 homo sapien

ALIGNMENTS

RESULT 1  
O96995 ID O96995 PRELIMINARY; PRT; 480 AA.  
AC O96995;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE NOTCHLESS protein.  
GN NLE OR NOTCHLESS OR CG2863.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99077802; PubMed=9857191;  
RA Royet J., Bouwmeester T., Cohen S.M.;  
RT "Notchless encodes a novel WD40-repeat-containing protein that modulates Notch signaling activity.";  
RL EMBO J. 17:7351-7360(1998).  
CC -!- SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).  
DR EMBL; AJ012588; CAA10070.1; -.  
DR FlyBase; FBgn0021874; Nle.  
DR InterPro; IPR001632; GproteIn\_B.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 8.  
DR PRINTS; PR00319; GPROTEINB.  
DR PRINTS; PR00320; GPROTEINBRPT.  
DR PRODOM; PD000018; WD40; 6.  
DR SMART; SM00320; WD40; 8.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_2.  
DR PROSITE; PS50082; WD\_REPEATS\_2; 7.  
DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
KW Repeat; WD repeat.  
SQ SEQUENCE 480 AA; 52933 MW; 40A5D696D33956C4 CRC64;

Query Match 99.7%; Score 2566; DB 5; Length 480;  
Best Local Similarity 99.8%; Pred. No. 4.5e-212;



RESULT 3		01-OCT-2000 (TrEMBLrel. 15, Created)	
Q8T4A2	Q8T4A2	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	DT
ID	Q8T4A2	01-MAR-2002 (TrEMBLrel. 21, Last sequence update)	DT
AC	Q8T4A2	CDNA FLJ10458 fis, clone NT2RP1001457, highly similar to Homo sapiens partial beta-transducin family protein (Hypothetical 53.3 kDa protein)	DE
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	partial beta-transducin family protein (Hypothetical 53.3 kDa protein)	DE
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	Homo sapiens (Human)	OS
DE	AT08344p.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	OC
GN	NLE.	NCBI_TaxID=9606;	OC
OS	Drosophila melanogaster (Fruit fly).	[1]	RN
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	SEQUENCE FROM N.A.	RP
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M., Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Takiguchi S., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Ninomiya K., Iwayanagi T.,	RA
OC	Ephydroidea; Drosophilidae; Drosophila.	"NEDO human cDNA sequencing project."	RA
OX	NCBI_TaxID=7227;	Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.	RL
RN	[1]	[2]	RN
RP	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	RP
RA	Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J., Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E., George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G., Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S., Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M., Celniker S.,	TISSUE=PLACENTA;	RC
RA	Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.	Strausberg R.;	RA
RL	EMBL; AY089286; AAL90024.1; -	Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.	RA
DR	EMBL; AY089286; AAL90024.1; -	- - SIMILARITY: CONTAINS 8 WD REPEATS (TRP-ASP DOMAINS).	RL
SQ	SEQUENCE 488 AA; 53809 MW; D4883202954A7785 CRC64;	EMBL; AK001320; BAA91621.1; -	CC
Query Match		DR EMBL; BC012075; AAL12075.1; -	DR
Best Local Similarity		InterPro; IPR001632; Gprotein_B.	DR
Matches 479; Conservative 0; Mismatches 1; Indels 1; Gaps 1;		InterPro; IPR001680; WD40.	DR
		Pfam; PF00400; WD40; 8.	DR
		PRINTS; PR00319; GPROTEINB.	DR
		PRINTS; PR00320; GPROTEINBPT.	DR
		ProDom; PD000018; WD40; 7.	DR
		SMART; SM00320; WD40; 8.	DR
		PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_4.	DR
		PROSITE; PS50082; WD_REPEATS_2; 7.	DR
		PROSITE; PS50294; WD_REPEATS_REGION; 1.	DR
		Hypothetical protein; Repeat; WD repeat.	DR
		SEQUENCE 485 AA; 53266 MW; 0AEBE24B44957379 CRC64;	KW
			SQ
		Query Match	
		Best Local Similarity	
		Matches 279; Conservative 73; Mismatches 118; Indels 6; Gaps 3;	
QY	9 EATPTIQARLVYTGEA----	9 EATPTIQARLVYTGEA----	QY
Db	8 EAVARDVQRLLVQFQDEGQLGSPFDVDPDITPDRLQLVCNALLAQEDPLPLAFFVHDA	8 EAVARDVQRLLVQFQDEGQLGSPFDVDPDITPDRLQLVCNALLAQEDPLPLAFFVHDA	Db
QY	65 EIKKSLEDTLDLASVDTENVIDIYQQAQVFKVRPVTRCTSSMPGHAEAVVSLNFPDGA	65 EIKKSLEDTLDLASVDTENVIDIYQQAQVFKVRPVTRCTSSMPGHAEAVVSLNFPDGA	QY
Db	68 EIVSSLGKLTESQAVETEKVLDIYQQAIFRVRVTRCTSSLEGHSEAVISVAFSPGK	68 EIVSSLGKLTESQAVETEKVLDIYQQAIFRVRVTRCTSSLEGHSEAVISVAFSPGK	Db
QY	125 HIASGSGDTTVRLWDLNLTETPHFTCTGKHQWVLCVSWAPDGRKRLASGCKAGSI	125 HIASGSGDTTVRLWDLNLTETPHFTCTGKHQWVLCVSWAPDGRKRLASGCKAGSI	QY
Db	128 YLASGSGDTTVRFWDLSTETPHFTCTGKHQWVLCVSWAPDGRKRLASGCKAGSI	128 YLASGSGDTTVRFWDLSTETPHFTCTGKHQWVLCVSWAPDGRKRLASGCKAGSI	Db
QY	185 GQOKGRPLSGHKHINCLAWEPYHRDPECKRLASASGDCGDCRIWDVKLGQCLMNIAGHTN	185 GQOKGRPLSGHKHINCLAWEPYHRDPECKRLASASGDCGDCRIWDVKLGQCLMNIAGHTN	QY
Db	188 GKQVGRTLAGHSKWITGLSWEPLHANPECRYVASSKDGSRVIRDTTACRCERILTGHTQ	188 GKQVGRTLAGHSKWITGLSWEPLHANPECRYVASSKDGSRVIRDTTACRCERILTGHTQ	Db
QY	245 AVTAVRWGGAGLIYTSKDRVTKMWRAADGILCRFTSGHAHWNINIALSTDYVLRGPH	245 AVTAVRWGGAGLIYTSKDRVTKMWRAADGILCRFTSGHAHWNINIALSTDYVLRGPH	QY
Db	248 SVTCLRWGGDGLYSASQDRTIKVWRAHDGVLCTLQGHGHWNTMALSTDYALRTGAFE	248 SVTCLRWGGDGLYSASQDRTIKVWRAHDGVLCTLQGHGHWNTMALSTDYALRTGAFE	Db
QY	305 PVK-DRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDNTLYLWRNNQN-KCVER	305 PVK-DRSKSHLSLSTEELQESALKRYQAVCPDEVESLVSCSDNTLYLWRNNQN-KCVER	QY
Db	308 PAEASVNPQDLQGLQELKERALSRYNLVRGQGERLVSGSDDFTLFLWSPAEDKKPLTR	308 PAEASVNPQDLQGLQELKERALSRYNLVRGQGERLVSGSDDFTLFLWSPAEDKKPLTR	Db
QY	363 MTGHQNVVNDVKYSPDVKLIASAFDCKSVRLWRASDGOYMATFRGHVQAVYTVAWSADR	363 MTGHQNVVNDVKYSPDVKLIASAFDCKSVRLWRASDGOYMATFRGHVQAVYTVAWSADR	QY
Db	368 MTGHQALINQVLFSPDSRIVASAFDCKSVRLWRASDGOYMATFRGHVQAVYTVAWSADR	368 MTGHQALINQVLFSPDSRIVASAFDCKSVRLWRASDGOYMATFRGHVQAVYTVAWSADR	Db











```
QY 374 KYSPDVKLLIASASFDKSVRLWRASDGOYMATFRGHVQAVYTVAWSADSRLLVSGSKDSTL 433
Db 367 VFSPDTRYIASASFDKSVKLCWCGRTGKYLASLRGHVGPYQVAWSADSRLLVSGSADSTL 426

QY 434 KWVSQVQTKLAQELPGHADEVFGVDWAPDGSRSVASGGKDKVIKWLW 478
Db 427 KVFEKTKSLYLDLPGHGDEVFTVDSPEGTKVVSGGKDKVLKLW 471

RESULT 11
O74855
ID O74855 PRELIMINARY; PRT; 502 AA.
AC O74855;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Notchless-like, WD repeat protein.
GN SPCC18.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Hilbert H., Duesterhoeft A., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 9 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL; AL031907; CAA21419.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 5.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_5.
DR PROSITE; PS50082; WD_REPEATS_2; 7.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 502 AA; 56003 MW; 5773208C7DEEB23A CRC64;
```

```
Query Match 40.7%; Score 1047.5; DB 3; Length 502;
Best Local Similarity 45.3%; Pred. No. 1.9e-81;
Matches 208; Conservative 76; Mismatches 160; Indels 15; Gaps 6;

QY 32 LPAGITTQQLGLICNALLKN-EEATPYLFFVGEDEIKKSLEDTDLIA-----SYDTEENVI 85
Db 49 VPGNSSVRQLEALLNQLLENSDDPVYPNFALHHEDETIEIQDNLVTSVFHNGLMKRTEDHL 108

QY 86 DIVYQPOAVFKVPVTRCTSSMPGHAAVVSLSNFSPPD-GAHLASGSDTTVRLWDLNLTET 144
Db 109 TLLYTPQAVFRVRAVTRCTASMGHDGTIIISAQFSPSTSSRLVTGSGDFTARLWDCDTQT 168

QY 145 PHFTCTGHKQWVLCVSWAPDGGKRLASGCKAGSIIWDPETGQOKRPLSGHKKHINCLAW 204
Db 169 PIATMKGHTNWVSCVAVAPDASIIATGSDMTNTRFWDPKKGSPIGDALRRHTKPIMALCW 228

QY 205 EPYHRDPECRK--LASASGDGDCRIWDVKLGQCLMNIAGHTNAVTVRWGGAGLIYTSSK 262
Db 229 OPLHLAPDSGPYLLASGSKDNTVRVWNVKLRTLTLLFTLSGHTAPITCVRWGGQNWIISSSY 288

QY 263 DRTVKMWRADGILCRFTSGHAHVNNIALSTDYVLRGTFPHVPKDRSKSHLSLSTEELEQ 322
Db 289 DKTIRIWDKDGKCLHLKGHAARVNHLSLSTEHVLRSGAYDHTDFKPKSF-----SDER 343

QY 323 ESALKRYQAVCPDEVESLVSCSDNTLYLW-RNNQNKCVERTMCHQNVNDVKYSPDVKL 381
Db 344 RKAKERYEACLKQSGERLVASDDLQLILWDPPQKSTKPIITKMHGKQKVVNHASFSPDGR 403

QY 382 IASASEDKSVRLWRASDGOYMATFRGHVQAVYTVAWSADSRLLVSGSKDSTLKWVSQTK 441
Db 404 IATASFDSSVRLWDGKTGKFLATLRGHVAAVYQCAWSTDSRLLVSSSQDITLKWVDVRSK 463
```

```
QY 442 KLAQELPGHADEVFGVDWAPDGSRSVASGGKDKVIKWLWAY 480
Db 464 KMKFDLPGHEDQVFAVDWSPDQQRVASGGADRAVRIWSH 502

RESULT 12
Q8X1P4
ID Q8X1P4 PRELIMINARY; PRT; 1356 AA.
AC Q8X1P4;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Beta transducin-like protein HET-E2C*4.
GN HET-E.
OS Podospora anserina.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Lasiosphaeriaceae; Podospora.
OX NCBI_TaxID=5145;
RN [1]
RP SEQUENCE FROM N.A.
RA Espagne E., Balhadere P., Begueret J., Turcq B.;
RT "A new family of WD40 proteins implicated in vegetative
RT incompatibility; evidence for a major role of WD40 repeat domain in
RT the specificity of het-e and het-d genes.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF323583; AAL37299.1; -.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 10.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 10.
DR SMART; SM00320; WD40; 10.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_10.
DR PROSITE; PS50082; WD_REPEATS_2; 10.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
SQ SEQUENCE 1356 AA; 149702 MW; 33148AF4A7B82826 CRC64;
```

```
Query Match 23.9%; Score 616; DB 3; Length 1356;
Best Local Similarity 29.4%; Pred. No. 1.1e-43;
Matches 150; Conservative 86; Mismatches 203; Indels 72; Gaps 10;

QY 13 HTIQARLVYTGEAEAGPPIDLPAGITTQQLGLICNALLKNEEATPY---LFFVGEDEIKKS 69
Db 763 HTIRGLRIAIVRDG-----YRFALSYRMIIIEKAPLOAYTSALVFAPDTSMIKK 810

QY 70 LEDTLDLASVDTENVIDIVYQPOAVFKVPVTRCTSSMPGHAAVVSLSNFSPPDGAHLASG 129
Db 811 IFKKEEPGWISTISVVEAEW-----NACTQTLEHGSSVLSVAFSPDQQRVASG 859

QY 130 SGDTTVRLWDLNLTETPHFTCTGHKQWVLCVSWAPDGGKRLASGCKAGSIIWDPETGQOKG 189
Db 860 SDDKTIKIWDTASGTGTQTLEHGGSVWSVAFSPDQQRVASGSDDKTIKIWDASGTCT- 918

QY 190 RPLSGHKKHINCLAWEPYHR-----DPE 212
Db 919 QTLEHGSSVLSVAFSPDQQRVASGSGDKTIKIWDTASGTGTQTLEHGGSVWSVAFSPD 978

QY 213 CRKLASASGDGDCRIWDVKLGQCLMNIAGHTNAVTVRWGGAG-LIYTSSKDRTVKMWR 271
Db 979 GQRVASGSGDKTIKIWDTASGTCTQTLEHGNSVWSVAFSPDQQRVASGSDDKTIKIWDT 1038

QY 272 ADGILCRFTSGHAHVNNIALSTD--YVLRGTFPHVPK--DRSKSHLSLSTEELEQESALK 327
Db 1039 ASGTCTQTLEHGGSVWSVAFSPDQQRVASGSDNTTIKIWDASGTCTQTLEHGSDSVWS 1098

QY 328 RYQAVCPDEVESLVSCSDNTLYLWRNNQNKCVERTMCHQNVNDVKYSPDVKLIASASF 387
Db 1099 --VAFSPDG-QRVASGSDHTIKIWDASGTCTQTLEHGSDSVWSVAFSPDQQRVASGSI 1155

QY 388 DKSRLWRASDGOYMATFRGHVQAVYTVAWSADSRLLVSGSKDSTLKWVSQTKKLAQEL 447
Db 1156 DGTIKIWDASGTCTQTLEHGGSVWSVAFSPDQQRVASGSDGTIKIWDASGTCTQTL 1215

QY 448 PGHADEVFGVDWAPDGSRSVASGGKDKVIKWLW 478
```











GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 03:21:05 ; Search time 3984 Seconds  
(without alignments)  
11359.152 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattcccaaaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*\*

- 1: gb\_ba:\*\*
- 2: gb\_htg:\*\*
- 3: gb\_in:\*\*
- 4: gb\_om:\*\*
- 5: gb\_ov:\*\*
- 6: gb\_pat:\*\*
- 7: gb\_ph:\*\*
- 8: gb\_pl:\*\*
- 9: gb\_pr:\*\*
- 10: gb\_ro:\*\*
- 11: gb\_sts:\*\*
- 12: gb\_sy:\*\*
- 13: gb\_un:\*\*
- 14: gb\_vi:\*\*
- 15: em\_ba:\*\*
- 16: em\_fun:\*\*
- 17: em\_hum:\*\*
- 18: em\_in:\*\*
- 19: em\_mu:\*\*
- 20: em\_om:\*\*
- 21: em\_or:\*\*
- 22: em\_ov:\*\*
- 23: em\_pat:\*\*
- 24: em\_ph:\*\*
- 25: em\_pl:\*\*
- 26: em\_ro:\*\*
- 27: em\_sts:\*\*
- 28: em\_un:\*\*
- 29: em\_vi:\*\*
- 30: em\_htg\_hum:\*\*
- 31: em\_htg\_inv:\*\*
- 32: em\_htg\_other:\*\*
- 33: em\_htg\_mus:\*\*
- 34: em\_htg\_pln:\*\*
- 35: em\_htg\_rod:\*\*
- 36: em\_htg\_mam:\*\*
- 37: em\_htg\_vrt:\*\*
- 38: em\_sy:\*\*
- 39: em\_htgo\_hum:\*\*
- 40: em\_htgo\_mus:\*\*
- 41: em\_htgo\_other:\*\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1555	100.0	1555	6	AX025192	AX025192 Sequence
2	1528.2	98.3	1663	3	AY089286	AY089286 Drosophil
3	1521	97.8	1521	3	DME012588	AJ012588 Drosophil
c 4	1285.8	82.7	85095	3	AC004573	AC004573 Drosophil
5	1285.8	82.7	142257	2	AC017491	AC017491 Drosophil
6	1285.8	82.7	182726	3	AC008002	AC008002 Drosophil
c 7	1285.8	82.7	302473	3	AE003589	AE003589 Drosophil
8	478	30.7	1859	6	AX135811	AX135811 Sequence
9	477.4	30.7	1853	9	AK001320	AK001320 Homo sapi
10	477.4	30.7	1854	6	AX077627	AX077627 Sequence
11	477.4	30.7	1870	9	BC002884	BC002884 Homo sapi
12	477.4	30.7	2593	9	BC012075	BC012075 Homo sapi
13	477.4	30.7	2638	6	AX405715	AX405715 Sequence
14	476.6	30.6	1784	10	BC018399	BC018399 Mus muscu
15	466.2	30.0	1928	6	AX135807	AX135807 Sequence
16	464.6	29.9	2250	5	AF069737	AF069737 Xenopus l
17	249.4	16.0	1170	6	AX135809	AX135809 Sequence
c 18	210.8	13.6	316613	8	SCCHR111	X59720 S.cerevisia
19	159.8	10.3	1563	6	AX488920	AX488920 Sequence
20	120.4	7.7	378	6	AX072820	AX072820 Sequence
c 21	115.8	7.4	40305	8	SPCC18	AL031907 S.pombe c
22	111.6	7.2	1373	9	HSAJ5257	AJ005257 Homo sapi
c 23	90.4	5.8	31787	3	AF106576	AF106576 Caenorhab
c 24	87.8	5.6	215786	9	AC022916	AC022916 Homo sapi
25	86.2	5.5	203769	2	AL603745	AL603745 Mus muscu
26	86.2	5.5	216655	2	AC120837	AC120837 Mus muscu
27	84.6	5.4	132348	10	CNS07Y0Z	AL713882 Mus muscu
c 28	82.8	5.3	334520	1	AP003588	AP003588 Nostoc sp
c 29	79.8	5.1	34131	2	AC096796	AC096796 Rattus no
30	79.8	5.1	178633	2	AC115671	AC115671 Rattus no
31	79.4	5.1	1844	9	BC026080	BC026080 Homo sapi
32	79.4	5.1	333550	1	AP003593	AP003593 Nostoc sp
33	78	5.0	4722	8	AF323585	AF323585 Podospora
34	77	5.0	4619	1	AF115313	AF115313 Thermomon
35	76.2	4.9	2740	9	HSM804229	AL832918 Homo sapi
36	76.2	4.9	81575	8	AB009055	AB009055 Arabidops
c 37	74.4	4.8	2311	3	AY113628	AY113628 Drosophil
38	73.6	4.7	5413	8	AF323583	AF323583 Podospora
39	72.8	4.7	1209	3	AF100287	AF100287 Trypanoso
40	72	4.6	5413	8	AF323582	AF323582 Podospora
41	72	4.6	5413	8	AF323584	AF323584 Podospora
c 42	72	4.6	10280	1	AE012949	AE012949 Chlorobiu
43	71	4.6	600	11	CNS06EEP	AL395015 T3 end of
44	70.6	4.5	5411	8	FLABTLP	L28125 Podospora a
45	68.4	4.4	2977	6	AX430450	AX430450 Sequence

ALIGNMENTS

RESULT 1  
AX025192  
LOCUS AX025192  
DEFINITION Sequence 9 from Patent WO0026364.  
ACCESSION AX025192  
VERSION AX025192.1 GI:10186913  
KEYWORDS  
SOURCE  
ORGANISM  
Drosophila sp.  
Drosophila sp.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1555)  
REFERENCE  
AUTHORS Cohen,S., Bouwmeester,A. and Royet,J.  
TITLE Regulator of notch signaling activity

AX025192 1555 bp DNA linear PAT 16-SEP-2000



This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

FEATURES

Location/Qualifiers  
1..1663  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/map="21C6-21C6"  
1..1663  
/gene="Nle"  
/note="alignment with genomic scaffold AE003589"  
/db\_xref="FLYBASE:FBgn0021874"  
100..1566  
/gene="Nle"  
/note="Longest ORF"  
/codon\_start=1  
/product="AT08344p"  
/protein\_id="AAL90024.1"  
/db\_xref="GI:19527819"  
/db\_xref="FLYBASE:FBgn0021874"  
/translation="MLAKKQKMQETDTEQEAHPHTIQARLVSDTGEAGPPIDLPA  
TTQQLGLICNALKNEEATPYLFVGEDEIKKSLDTLDLASVDTENVIDIYQPAV  
FKVRPVTRCTSSMPGHAEEAVVSLNFSPDGAHLASGSDTIVRLWDLNTEPHETCTGH  
KQVLCVSWAPDKRLASGCKAGSIIWDPETGQKGRPLSGHKKHINCLAWEPYHRD  
PECKLASASGDGCRIDWDVILGQCLMNIAGHTNAVTVRWGAGLIYTSKDRTVKM  
WRAADGILCRTFESHAHWNNTALSTDYVLRTPGFPHVVKDRSKSHLSLSTEELQESAL  
KRYQAVCPDEVESLVSCDDNTLYLWRNNQKCVERMTHONVNDVKYSPDVKLIAS  
ASFDKSVRLWRASDQYMATFRGHVQAVYTVASADSRLIYVSGSKDSTLKVWSVQTKK  
LAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKLWAY"

gene

CDS

BASE COUNT  
ORIGIN

Query Match 98.3%; Score 1528.2; DB 3; Length 1663;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1543; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY	7	CAAAAAATGAGGAGACGGACACGGAGCAAGAGGCCACCCACATACGATACAGGCGCGC	66
Db	115	CAGAAAATGAGGAGACGGACACGGAGCAAGAGGCCACCCACATACGATACAGGCGCGC	174
QY	67	CTCGTTT---ACACGGGAGGAAGCCGCCCGCCCAATCGACCTGCCGGCAGGAATCACT	123
Db	175	CTCGTTTCGGACACGGGCGAGGAAGCCGCCCGCCCAATCGACCTGCCGGCAGGAATCACT	234
QY	124	ACCCAGCAATTGGGACTGATTTGCAACCGCGTCTGTGAAAAACGAGGAGCCACCTCCATAT	183
	235	ACCCAGCAATTGGGACTGATTTGCAACCGCGTCTGTGAAAAACGAGGAGCCACCTCCATAT	294
	184	TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCGCTGGAGGACAGCTTGGACTTGGCG	243
		TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCGCTGGAGGACAGCTTGGACTTGGCG	354
		STGGACACCGAAACCGTGATCGATATTGTGTATCAGCCACAGCGGTTTTCAAAGTG	303
		CACCGAAACCGTGATCGATATTGTGTATCAGCCACAGCGGTTTTCAAAGTG	414
		GATGCACGAGTTCCATGTCGGGACACCGCGAGCGCTGTGTTTCGCTG	363
		ACGAGTTCCATGTCGGGACACCGCGAGCGCTGTGTTTCGCTG	474
		TCGCGCCAGTGGAAAGTGGCGACACCAACAGTGGGATTG	423

Db	475	AATTTACGCCCCGATGGTGCTCATCTCGCCAGTGGAGTGGCGACACACAGTGGGATTG	534
QY	424	TGGGATCTTAACACAGACACACCGCACTTACCTGCACAGGTCTAAGCAGTGGTCTG	483
Db	535	TGGGATCTTAACACAGACACACCGCACTTACCTGCACAGGTCTAAGCAGTGGTCTG	594
QY	484	TGCGTATCTTGGCTCCGGATGGCAACGGTTGGCCAGCGTTGCAAGCGGGCTCTATA	543
Db	595	TGCGTATCTTGGCTCCGGATGGCAACGGTTGGCCAGCGTTGCAAGCGGGCTCTATA	654
QY	544	ATCATCTGGGACCCCGAGACGGGTACAGAGAAGGGGCGACCTTGCAGTGGGCAAGAAA	603
Db	655	ATCATCTGGGACCCCGAGACGGGTACAGAGAAGGGGCGACCTTGCAGTGGGCAAGAAA	714
QY	604	CACATCAACTGCCTCGCCTGGGAACCGTATCATCCGGATCCGGAGTGCAGGAACTTGCT	663
Db	715	CACATCAACTGCCTCGCCTGGGAACCGTATCATCCGGATCCGGAGTGCAGGAACTTGCT	774
QY	664	TCCGCCAGTGGAGACGGGGACTGCCGGATTGGGACGTAATAATTGGGCCAGTGCCTTATG	723
Db	775	TCCGCCAGTGGAGACGGGGACTGCCGGATTGGGACGTAATAATTGGGCCAGTGCCTTATG	834
QY	724	AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGCGGGCTTATT	783
Db	835	AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGCGGGCTTATT	894
QY	784	TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGTGATGGAATCTTGTC	843
Db	895	TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGTGATGGAATCTTGTC	954
QY	844	CGGAGCTTCTTGCCAAAGCTCACTGGGTAAACAACATTTGCCGTGAGCACCGGATTACGTC	903
Db	955	CGGAGCTTCTTGCCACGCTCACTGGGTAAACAACATTTGCCGTGAGCACCGGATTACGTC	1014
QY	904	CTGGCAGTGGTCCATCCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC	963
Db	1015	CTGGCAGTGGTCCATCCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC	1074
QY	964	ACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACAGGCGCTGTCCTGACGAGGTG	1023
Db	1075	ACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACAGGCGCTGTCCTGACGAGGTG	1134
QY	1024	GAGTCGCTGTTTCTGTTCCGATGACAACACCTCTATCTGTGGCGGAACACCAAGAAC	1083
Db	1135	GAGTCGCTGTTTCTGTTCCGATGACAACACCTCTATCTGTGGCGGAACACCAAGAAC	1194
QY	1084	AAGTCGTTGAGCGCATGACAGGCGCACCAAGAGTGGTCAACGATGTGAATATTTCGCCG	1143
Db	1195	AAGTCGTTGAGCGCATGACAGGCGCACCAAGAGTGGTCAACGATGTGAATATTTCGCCG	1254
QY	1144	GATGTAAGCTAAATTCGCTCTGCTTCAATTTGACAAAGTCAAGTGGTGTGGGAGCCAGC	1203
Db	1255	GATGTAAGCTAAATTCGCTCTGCTTCAATTTGACAAAGTCAAGTGGTGTGGGAGCCAGC	1314
QY	1204	GATGTCAGTACATGGCCACCTTCCGGGGTCAATGTGCAGGCTGTACACGTTGCCTGG	1263
Db	1315	GATGTCAGTACATGGCCACCTTCCGGGGTCAATGTGCAGGCTGTACACGTTGCCTGG	1374
QY	1264	TCCGGCGACTCCCGCTTGATGTTTCCGGGACCAAGACTCAACTCTAAAAGTATGGAGT	1323
Db	1375	TCCGGCGACTCCCGCTTGATGTTTCCGGGACCAAGACTCAACTCTAAAAGTATGGAGT	1434
QY	1324	GTGCAGACGAAGAACTGGCACAGGAGCTGCCTGGACATGCCGATGAGGTTCGGAGTG	1383
Db	1435	GTGCAGACGAAGAACTGGCACAGGAGCTGCCTGGACATGCCGATGAGGTTCGGAGTG	1494
QY	1384	GACTGGGCGCGCATGGCTTAGAGTTGCCTCTGGTGGCAAGGACAAAAGTATATAAGCTA	1443
Db	1495	GACTGGGCGCGCATGGCTTAGAGTTGCCTCTGGTGGCAAGGACAAAAGTATATAAGCTA	1554
QY	1444	TGGGCTTATTAAACAAATCAATTAACCTGTACACGGTAAGAAAATACCTTAGGAATAAAGTAA	1503



ACGGTAAGAAAAATACTTAGGAATAAAGTAA 1614  
AAAAAATAAAAAAATAAAAAA 1552  
AAAAAATAAAAAAATAAAAAA 1663

1 bp mRNA linear INV 12-MAR-1999  
NA for Notchless protein.

ats.

hropoda; Hexapoda; Insecta; Pterygota;  
a; Diptera; Brachycera; Muscomorpha;  
idae; Drosophila.

REFERENCE Royet J., Developmental Biology Programme,  
AUTHORS Direct Subm. 998) Royet J., Developmental Biology Programme,  
TITLE EMBL, Meyerhofstrasse 1 Heidelberg, 69117, GERMANY  
JOURNAL 2 (bases 1 to 1521)  
REFERENCE Royet, J., Bouwmeester, T. and Cohen, S.M.  
AUTHORS Notchless encodes a novel WD40-repeat-containing protein that  
TITLE modulates Notch signaling activity  
JOURNAL EMBO J. 17 (24), 7351-7360 (1998)  
MEDLINE 99077802  
PUBMED 9857191

FEATURES  
source Location/Qualifiers  
1..1521  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
1..1443  
/gene="notchless"  
1..1443  
/gene="notchless"  
/note="WD40 repeat containing protein"  
/codon\_start=1  
/product="Notchless protein"  
/protein\_id="CAA10070.1"  
/db\_xref="GI:4127781"

gene  
CDS

/translation="MQETDTEQATPHTIQARLVYTGEEAGPPIDLPAGITQQGLI  
CNALLKNEATPYLFFVGEDEIKKSLDTLDASVDTEWVIVQPAVFKVRPVR  
CTSMMPGHAEEAVSLNFSPDGAHLASGSDTVRLWDLNTEPHFTCTGKHQWVLCVS  
WAPGKRLASGKAGSIIWDPETGQKGRPLSGHKKHINCLAWEPYHPRPECKLAS  
ASGDCRDIWDVKLGCLMNIAGHTNAVTAVMGGAGLIYTSKDRTVKMWRAADGIL  
CRFESQAHWNNTALSTDYVLTGPFHPVDRSKSHLSLSTEELOESALKRYQAVCP  
DEVESLVSCDDNTLYLRNNQNKVERMTGHQNVNDVYSPDVVKLIASASFDKSVR  
LWRASDQYMAFERGHVQAVYVAVWSADSRLLVSGSKDSLKWSVQTKLAQELPGH  
ADEVFQVDWAPDGSRVASGGKDKVIKLWAY"

BASE COUNT 389 a 370 c 448 g 314 t  
ORIGIN

Query Match 97.8%; Score 1521; DB 3; Length 1521;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ATGCAGGAGACGGACACGGAGCAAGAGCGCCACATACGATACAGCGCGCCTCGTT 72  
Db 1 ATGCAGGAGACGGACACGGAGCAAGAGCGCCACATACGATACAGCGCGCCTCGTT 60  
QY 73 TACACGGCGGAGGAAGCGCGCCCAATCGACCTGCCGCGAGGAATCACTACCCAGCAA 132  
Db 61 TACACGGCGGAGGAAGCGCGCCCAATCGACCTGCCGCGAGGAATCACTACCCAGCAA 120  
QY 133 TTGGGACTGATTGCAACGGCGCTGCTGAAACAGGAGGAGCCACTCCATATTGTTTTTC 192  
Db 121 TTGGGACTGATTGCAACGGCGCTGCTGAAACAGGAGGAGCCACTCCATATTGTTTTTC 180  
QY 193 GTGGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCGTCAGTGGAC 252

Db 181 GTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCGTCAGTGGAC 240  
QY 253 ACCGAAAACTGATCGATATGTGTATCAGCCACAGCGGTTTTTCAAAGTGGCCCGAGTG 312  
Db 241 ACCGAAACGTGATCGATATGTGTATCAGCCACAGCGGTTTTTCAAAGTGGCCCGAGTG 300  
QY 313 ACAAGATGCACGAGTTCATGCCCGGGACAGCCGAGGCTGTGGTTTCGCTGAATTTTCAGC 372  
Db 301 ACAAGATGCACGAGTTCATGCCCGGGACAGCCGAGGCTGTGGTTTCGCTGAATTTTCAGC 360  
QY 373 CCGGATGGTCTCATCTCGCCAGTGGAGTGGCGGACACACAGTGGGATCTT 432  
Db 361 CCGGATGGTCTCATCTCGCCAGTGGAGTGGCGGACACACAGTGGGATCTT 420  
QY 433 AACACAGAGACACCGCACCTTCACCTGCACAGGTCTAAGCAGTGGGTTCTGTGCGTATCC 492  
Db 421 AACACAGAGACACCGCACCTTCACCTGCACAGGTCTAAGCAGTGGGTTCTGTGCGTATCC 480  
QY 493 TGGGCTCCGGATGGCAACCGTTGGCCAGCGGTTGCAAGCGGGCTCTATAATCATCTGG 552  
Db 481 TGGGCTCCGGATGGCAACCGTTGGCCAGCGGTTGCAAGCGGGCTCTATAATCATCTGG 540  
QY 553 GACCCGGAGACGGGTGACGACAGAGGGGGCGGCCCTTGTAGTGGGCACAAAGAACACATCAAC 612  
Db 541 GACCCGGAGACGGGTGACGACAGAGGGGGCGGCCCTTGTAGTGGGCACAAAGAACACATCAAC 600  
QY 613 TGCCTCGCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGTCTCCGCCAGT 672  
Db 601 TGCCTCGCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGTCTCCGCCAGT 660  
QY 673 GGAGACGGGGACTGCCGGATTGGGACGTAATAATTGGGCCAGTGCCTTATGAACATTTGCC 732  
Db 661 GGAGACGGGGACTGCCGGATTGGGACGTAATAATTGGGCCAGTGCCTTATGAACATTTGCC 720  
QY 733 GGACACACAAATGCTGTGACACAGTGTGGGTGGAGCGGGCTTATTTATATACATCC 792  
Db 721 GGACACACAAATGCTGTGACACAGTGTGGGTGGAGCGGGCTTATTTATATACATCC 780  
QY 793 TCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGCGGACGTTTC 852  
Db 781 TCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGCGGACGTTTC 840  
QY 853 TCTGGCCAAAGCTCACTGGGTAAACAACATTTGGCTGAGCACCGGATTACGCTCGCGCACT 912  
Db 841 TCTGGCCAAAGCTCACTGGGTAAACAACATTTGGCTGAGCACCGGATTACGCTCGCGCACT 900  
QY 913 GGTCCATTCCATCCCGGTGAAGATCGCTCCAAAGAGCCACCTCAGTTTGTAGCACTGAGGAA 972  
Db 901 GGTCCATTCCATCCCGGTGAAGATCGCTCCAAAGAGCCACCTCAGTTTGTAGCACTGAGGAA 960  
QY 973 TTGCAGGAATCTGCCTTGAAGCGCTACCAGCGCGTGTGCCCTGACGAGGTGGAGTCGCTG 1032  
Db 961 TTGCAGGAATCTGCCTTGAAGCGCTACCAGCGCGTGTGCCCTGACGAGGTGGAGTCGCTG 1020  
QY 1033 GTTTCCTGTTCCGATGACAAACCCCTCTATCTGTGGCGGAACAACCAAGTGCCTT 1092  
Db 1021 GTTTCCTGTTCCGATGACAAACCCCTCTATCTGTGGCGGAACAACCAAGTGCCTT 1080  
QY 1093 GAGCGCATGACAGGGCACCAAGACGTGGTCAACGATGTGAATATTCGCCGATGTAAAG 1152  
Db 1081 GAGCGCATGACAGGGCACCAAGACGTGGTCAACGATGTGAATATTCGCCGATGTAAAG 1140  
QY 1153 CTAATTCGCTCTGCTTCATTTGACAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1212  
Db 1141 CTAATTCGCTCTGCTTCATTTGACAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1200  
QY 1213 TACATGGCCACCTTCCGGGTCTATGTGAGGTGTGTTTACAGGTGCTTACAGGTGCTTCCGCGGAC 1272  
Db 1201 TACATGGCCACCTTCCGGGTCTATGTGAGGTGTGTTTACAGGTGCTTCCGCGGAC 1260  
QY 1273 TCCGCTTGATTGTTCCGGCAGCAAGACTCAACTCTAAAGTATGGAGTGTGCAGACG 1332



Db 1261 TCCCGCTGATGTTTCCGGCAGCAAGACTCAACTCTAAAAGTATGGAGTGTGCAGACG 1320

QY 1333 AAGAACTGGCACAGGAGCTGCCTGGACATCGCGATGAGGTGTTTCGGAGTGGACTGGCG 1392

Db 1321 AAGAACTGGCACAGGAGCTGCCTGGACATCGCGATGAGGTGTTTCGGAGTGGACTGGCG 1380

QY 1393 CCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGACAAAAGTTATAAAGCTATGGGCTTAT 1452

Db 1381 CCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGACAAAAGTTATAAAGCTATGGGCTTAT 1440

QY 1453 TAACAAATCATTAACCTGTACACGGTAAGAAATACTTAGGAATAAAGTAAACGTCCTG 1512

Db 1441 TAACAAATCATTAACCTGTACACGGTAAGAAATACTTAGGAATAAAGTAAACGTCCTG 1500

QY 1513 AGTAAAAAATAAATAAATAAATAA 1533

Db 1501 AGTAAAAAATAAATAAATAAATAA 1521

RESULT 4  
AC004573/c

LOCUS  
Drosophila melanogaster, chromosome 2L, region 21C5-21D1, pl clone

DEFINITION  
DS07610, complete sequence.

ACCESSION  
AC004573

VERSION  
AC004573.1 GI:4204255

KEYWORDS  
HTG.

SOURCE  
Drosophila melanogaster.

ORGANISM  
Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
1 (bases 1 to 85095)

AUTHORS  
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

TITLE  
Sequencing of Drosophila chromosome 2L, region 21C5-21D1

JOURNAL  
Unpublished (1998)

REFERENCE  
2 (bases 1 to 85095)

AUTHORS  
Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.

TITLE  
Direct Submission

JOURNAL  
Submitted (17-APR-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT  
On Jan 30, 1999 this sequence version replaced gi:4164118.  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory, MS 64-121  
Berkeley, CA 94720

For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgpefruitfly.berkeley.edu](mailto:bdgpefruitfly.berkeley.edu).

P1 library location: 80-26.  
Location/Qualifiers  
1. .85095  
/organism="Drosophila melanogaster"  
/strain="y2; cn bw sp"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
/map="21C5-21D1"

/clone="pl.DS07610 (D275)"  
/clone\_lib="Pl library, partial Sau3A in pNS582tet14ad10"  
BASE COUNT 22289 a 19671 c 20319 g 23816 t  
ORIGIN

Query Match 82.7%; Score 1285.8; DB 3; Length 85095;  
Best Local Similarity 89.2%; Pred. No. 5.5e-311;  
Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;

QY 7 CAAAAATGCAGGAGACGACACGGAGCAAGAGGCCACCATACATACAGCGCGC 66

Db 27639 CAGAAAATGCAGGAGACGACACGGAGCAAGAGGCCACCATACAGCGCGC 27580

QY 67 CTCGTTT---ACACGGCGAGGAAGCCGGCCCGCCCAATCGACCTGCCGGCAGGAATCACT 123

Db 27579 CTCGTTTCGGACACGGCGAGGAAGCGGGCCCGCCAATCGACCTGCCGGCAGGAATCACT 27520

QY 124 ACCAGCAATTTGGACTGATTTGCAACGCGCTGTGTAAGAAACGAGGAAGCCACTCCATAT 183

Db 27519 ACCAGCAATTTGGTCTGATTTGCAACGCGCTGTGTAAGAAACGAGGAAGCCACTCCATAT 27460

QY 184 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 243

Db 27459 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 27400

QY 244 TCAGTGGACACCGAAACGTCGATTCGATATTGTGTATCAGCCACAGCGGGTTTCAAAGTG 303

Db 27399 TCAGTGGACACCGAAACGTCGATTCGATATTGTGTATCAGCCACAGCGGGTTTCAAAGTG 27340

QY 304 CGCCAGTGACAAGATGCACGAGTTCCTATCCCGGACACGCCGAGGCTGTGGTTTCGCTG 363

Db 27339 CGCCAGTGACAAGATGCACGAGTTCCTATCCCGGACACGCCGAGGCTGTGGTTTCGCTG 27280

QY 364 AATTCAGCCCGGATGGTGTCTATCTCGCCAGTGGAAAGTGGCGACACACAGTGGGATTG 423

Db 27279 AATTCAGCCCGGATGGTGTCTATCTCGCCAGTGGAAAGTGGCGACACACAGTGGGATTG 27220

QY 424 TGGGATCTTAACACAGAGACACCCGACTTCACCTGCACAGGTATATAAGCAGTGGTTCTG 483

Db 27219 TGGGATCTTAACACAGAGACACCCGACTTCACCTGCACAGGTATATAAGCAGTGGTTCTG 27160

QY 484 TGCCTATCTCTGGCTCCGGATCGGATGSGAAACGGTTGGCCAGCGGTTGCAAGCGGCTCTATA 543

Db 27159 TGCCTATCTCTGGCTCCGGATGSGAAACGGTTGGCCAGCGGTTGCAAGCGGCTCTATA 27100

QY 544 ATCATCTGGACCCGGAGACGGGTTCAGCAGAAAGGGCGCACCTTGAGTGGGCACAAGAAA 603

Db 27099 ATCATCTGGACCCGGAGACGGGTTCAGCAGAAAGGGCGCACCTTGAGTGGGCACAAGAAA 27040

QY 604 CACATCAACTGCCTGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAACCTTGCT 663

Db 27039 CACATCAACTGCCTGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAACCTTGCT 26980

QY 664 TCCGCCAGTGGAGACGGGACTGCCGGATTTCGGACGTAAATTTGGCCAGTGCCTTATG 723

Db 26979 TCCGCCAGTGGAGACGGGACTGCCGGATTTCGGACGTAAATTTGGCCAGTGCCTTATG 26920

QY 724 AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGCGGCCCTTATT 783

Db 26919 AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGCGGCCCTTATT 26860

QY 784 TATACATCTCCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC 843

Db 26859 TATACATCTCCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC 26800

QY 844 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATTTGGCTGAGCACCCGATTACGTC 903

Db 26799 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATTTGGCTGAGCACCCGACTACGTC 26740

QY 904 CTGCGCACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC 963

Db 26739 CTGCGCACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC 26680

```
QY 964 A-----C 965
|
Db 26679 AGTAAGAAATTCATATGCTCTCTATATATGCTGTATATCCTAAAGAGATTGTTTACAGC 26620
|
QY 966 TGAGGAATGCAGGAATCTGCCCTTGAAAGCGCTACCAAGCCCGTGTGCCCTGACGAGGTGGA 1025
|
Db 26619 TGAGGAATGCAGGAATCTGCCCTTGAAAGCGCTACCAAGCCCGTGTGCCCTGACGAGGTGGA 26560
|
QY 1026 GTCGCTGGTTCCCTGTTCCGATGACACACCCCTCTATCTGTGGCGGAACAACCAACAACAA 1085
|
Db 26559 GTCGCTGGTTCCCTGTTCCGATGACACACCCCTCTATCTGTGGCGGAACAACCAACAACAA 26500
|
QY 1086 GTGCGTTGAGCGCATGACAGGGCACCAGAACGTGGTCAACGATGTGAAATATTCGCCGGA 1145
|
Db 26499 GTGCGTTGAGCGCATGACAGGGCACCAGAACGTGGTCAACGACGTGAAATATTCGCCGGA 26440
|
QY 1146 TGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTCCGTCTGTGGCGAGCCAGCGA 1205
|
Db 26439 TGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTCCGTCTGTGGCGAGCCAGCGA 26380
|
QY 1206 TGGTCAGTACATGGCCACCTTCCGGGGTCATGTGCAGGCTGTTTACACGGTGTGCCCTGGTC 1265
|
Db 26379 TGGTCAGTACATGGCCACCTTCCGGGGTCATGTGCAGGCTGTTTACACGGTGTGCCCTGGTC 26320
|
QY 1266 CGCGGACTCCCGCTTGATTGTTTCCGGCAGCAAGACTCAAACTCTAAA----- 1313
|
Db 26319 CGCGGACTCCCGCTTGATTGTTTCCGGCAGCAAGACTCAAACTCTAAAAGGTTAGCACGT 26260
|
QY 1314 -----AGTATGGAGTGTGCAGA 1330
|
Db 26259 AGTTTCAGTTTCTAGCGCTAGCTAATACATATCCTCTTGCTTAGTATGGAGTGTGCAGA 26200
|
QY 1331 CGAAGAAACTGGCACAGGAGCTGCCCTGGACATCGGGATGAGGTGTTCCGGAGTGGACTGGG 1390
|
Db 26199 CGAAGAAACTGGCACAGGAGCTGCCCTGGACATCGGGATGAGGTGTTCCGGAGTGGACTGGG 26140
|
QY 1391 CGCCCGATGGCTCTAGAGTTGCCTCTGGTGCAAGGACAAAAGTTATAAGCT----- 1442
|
Db 26139 CGCCCGATGGCTCTAGAGTTGCCTCTGGTGCAAGGACAAAAGTTATAAGCTGTAAGTTT 26080
|
QY 1443 -----ATGGGCTTA 1451
|
Db 26079 GAGTAACGAGGTTTCTATAACAATATACAGACTCTAATTGTATCTTTTCAGATGGGCTTA 26020
|
QY 1452 TTAACAAATCATTAACCTTGTCACGGGTAGAAAATACTTAGGAATAAAGTAAACGTCCT 1511
|
Db 26019 TTAACAAATCATTAACCTTGTCACGGGTAGAAAATACTTAGGAATAAAGTAAACGTCCT 25960
|
QY 1512 GAGTAAA 1518
|
Db 25959 GAGTAAA 25953
```

```
RESULT 5
AC017491
LOCUS Drosophila melanogaster, 142257 bp DNA linear HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017491
VERSION AC017491.1 GI:6553495
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 142257)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210988 by the submitter.
For more information on this record e-mail to fly@celera.com.
```

\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
    source  
        Location/Qualifiers  
            1. .142257  
            /organism="Drosophila melanogaster"  
            /db\_xref="taxon:7227"

BASE COUNT 38086 a 33712 c 32464 g 37995 t  
ORIGIN

Query Match 82.7%; Score 1285.8; DB 2; Length 142257;  
Best Local Similarity 89.2%; Pred. No. 5.9e-311;  
Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;

```
QY 7 CAAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCGC 66
|
Db 100034 CAGAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCGC 100093
|
QY 67 CTCGTTT---ACACGGGGCGAGGAAGCCGGCCGCCCAATCGACCTGCCGGCAGGAATCACT 123
|
Db 100094 CTCGTTTCGGACACGGGGCGAGGAAGCGGGCCGCCAATCGACCTGCCGGCAGGAATCACT 100153
|
QY 124 ACCCAGCAATTGGGACTGATTGTCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 183
|
Db 100154 ACCCAGCAATTGGGTCTGATTGTCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 100213
|
QY 184 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTTTGACTTGGCG 243
|
Db 100214 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTTTGACTTGGCG 100273
|
QY 244 TCAGTGGACACCGAAAAACGTGATCGATATTGTGTATCAGCCACAGGCGGTTTCAAAGTG 303
|
Db 100274 TCAGTGGACACCGAAAAACGTGATCGATATTGTGTATCAGCCACAGGCGGTTTCAAAGTG 100333
|
QY 304 CGCCAGTGACAAGATGCACGAGTTCATTCGCGGGACACGCCGAGGCTGTGGTTTCGCTG 363
|
Db 100334 CGCCAGTGACAAGATGCACGAGTTCATTCGCGGGACACGCCGAGGCTGTGGTTTCGCTG 100393
|
QY 364 AATTTCAGCCCGGATGGTGTCTATCTCGCCAGTGGGAAGTGGCGACACCCACAGTCGCGATTG 423
|
Db 100394 AATTTCAGCCCGGATGGTGTCTATCTCGCCAGTGGGAAGTGGCGACACCCACAGTCGCGATTG 100453
|
QY 424 TGGATCTTAACACAGAGACACCCGCACTTCACCTGCACAGGTCTATAAGCAGTGGGTTCTG 483
|
Db 100454 TGGATCTTAACACAGAGACACCCGCACTTCACCTGCACAGGTCTATAAGCAGTGGGTTCTG 100513
|
QY 484 TCGTATCCTGGGCTCCGGATGGCAAAACGGTTGGCCACGCGGTTGCAAAGCGGCTCTATA 543
|
Db 100514 TCGTATCCTGGGCTCCGGATGGCAAAACGGTTGGCCACGCGGTTGCAAAGCGGCTCTATA 100573
|
QY 544 ATCATCTGGGACCCGGAGACGGGTTCAGCAGAAGGGCGGACCCCTTGAGTGGGCACAAGAAA 603
|
Db 100574 ATCATCTGGGACCCGGAGACGGGTTCAGCAGAAGGGCGGACCCCTTGAGTGGGCACAAGAAA 100633
|
QY 604 CACATCAACTGCCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGCT 663
|
Db 100634 CACATCAACTGCCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGCT 100693
|
QY 664 TCCGCCAGTGGAGACGGGGACTGCCGGATTGCGGACGTAAAAATTGGGCCAGTGCCTTATG 723
|
Db 100694 TCCGCCAGTGGAGACGGGGACTGCCGGATTGCGGACGTGAAATTTGGGCCAGTGCCTTATG 100753
|
QY 724 AACATTGCCGGACACACAAAATGCTGTGACAGCAGTGAGATGGGGTGGAGCGGCGCTTATT 783
|
Db 100754 AACATTGCCGGACACACAAAATGCTGTGACAGCAGTGAGATGGGGTGGAGCGGCGCTTATT 100813
|
QY 784 TATACATCCTCCAAGATCGCACAGTGAAGATGTGCGGAGCAGCTGATGGAATCTTGTC 843
|
Db 100814 TATACATCCTCCAAGATCGCACAGTGAAGATGTGCGGAGCAGCTGATGGAATCTTGTC 100873
|
QY 844 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATTCGCTGAGACCCGATTACGTC 903
|
```

Db 100874 CGGACGTTCTCTGGCCACGCTCACTGGGTAACAACATTGCCCTGAGCACCGACTACGTC 100933

QY 904 CTGGCGACTGGTCCATTCATCCGGTGAAGGATCGCTCCAGAGCCACCTCAGTTTGAGC 963

Db 100934 CTGGCGACTGGTCCATTCATCCGGTGAAGGATCGCTCCAGAGCCACCTCAGTTTGAGC 100993

QY 964 A-----C 965

Db 100994 AGTAAGAATTTCATATGCTCTCTATATATGCTGTATATCCTCTATAAGAGATTCTTTTACAGC 101053

QY 966 TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTPACAGGCGCTGTGCCCTGACGAGGTGGA 1025

Db 101054 TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTPACAGGCGCTGTGCCCTGACGAGGTGGA 101113

QY 1026 GTCGCTGGTTTCTGTTCCGGATGACAACACCCCTCTATCTGTGGCGGAACCAACCAGAACAA 1085

Db 101114 GTCGCTGGTTTCTGTTCCGGATGACAACACCCCTCTATCTGTGGCGGAACCAACCAGAACAA 101173

QY 1086 GTGCGTTGAGCGCATGACAGGCGCACAGACGTGGTCAACGATGTGAATATTTCGCCGGA 1145

Db 101174 GTGCGTTGAGCGCATGACAGGCGCACAGACGTGGTCAACGACGTGAATATTTCGCCGGA 101233

QY 1146 TGTAAAGCTAATGTCGCTCTGCTTCATTTGACAAGTCAGTGGCTGTGTGGCGAGCCAGCGA 1205

Db 101234 TGTAAAGCTAATGTCGCTCTGCTTCATTTGACAAGTCAGTGGCTGTGTGGCGAGCCAGCGA 101293

QY 1206 TGGTCAGTACATGGCCACCTTCGGGGTCTATGTGCAGGCTGTTTACACGGTTGCCCTGGTC 1265

Db 101294 TGGTCAGTACATGGCCACCTTCGGGGTCTATGTGCAGGCTGTTTACACGGTTGCCCTGGTC 101353

QY 1266 CGCGGACTCCCGCTTGATTGTTTCCGGCAGCAAGACTCAACTCTAAA----- 1313

Db 101354 CGCGGACTCCCGCTTGATTGTTTCCGGCAGCAAGACTCAACTCTAAAAGGTTAGCACGT 101413

QY 1314 -----AGTATGGAGTGTGCAGA 1330

Db 101414 AGTTTCAGTTTCTAGCGCTAGCTAATACATATCCTCTGCTTAGTATGGAGTGTGCAGA 101473

QY 1331 CGAAGAACTGGCCACAGGAGCTGCCTGGACATCGGGATGAGGTGTTCCGAGTGGACTGGG 1390

Db 101474 CGAAGAACTGGCCACAGGAGCTGCCTGGACATCGGGATGAGGTGTTCCGAGTGGACTGGG 101533

QY 1391 CGCCCGATGGCTCTAGAGTGGCTCTGTGGCAAGGACAAAGTTATAAAGCT----- 1442

Db 101534 CGCCCGATGGCTCTAGAGTGGCTCTGTGGCAAGGACAAAGTTATAAAGCTGTAAAGTTT 101593

QY 1443 -----ATGGGCTTA 1451

Db 101594 GAGTAACGAGGTTTTCATATAACAATATACAGACTCTAATTGTATCTTTCAGATGGGCTTA 101653

QY 1452 TTAACAAATCATTAACCTGTACACGGTAAGAAAATACTTAGGAATAAGTAAAAACGTCCT 1511

Db 101654 TTAACAAATCATTAACCTGTACACGGTAAGAAAATACTTAGGAATAAGTAAAAACGTCCT 101713

QY 1512 GAGTAAA 1518

Db 101714 GAGTAAA 101720

RESULT 6

AC008002

LOCUS

DEFINITION

AC008002

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AC008002 182726 bp DNA linear INV 06-APR-2001

Drosophila melanogaster, chromosome 2L, region 21D-21E, BAC clone

BACR48E08, complete sequence.

GI:13559545

Drosophila melanogaster.

Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 182726)

AUTHORS

Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champs, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferrera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattel, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 182726)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Butenhoff, C., Champs, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

Direct Submission

Submitted (09-JUL-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Apr 6, 2001 this sequence version replaced gi:6532012.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1..182726

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="2L"

/map="21D-21E"

/clone="BACR48E08 (D843)"

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

PBACe3.6)"

BASE COUNT 49900 a 42515 c 41297 g 49014 t

ORIGIN

Query Match 82.7%; Score 1285.8; DB 3; Length 182726;

Best Local Similarity 89.2%; Pred. No. 6.1e-311;

Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;

QY 7 CAAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCACATACGATACAGGCGCGC 66

Db 89258 CAGAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCACATACGATACAGGCGCGC 89317

QY 67 CTCGTTT---ACACGGCGGAGGAAGCCGCCGCCCAATCGACCTGCCGCGAGGAATCACT 123

Db 89318 CTCGTTTCGGACACGGCGGAGGAAGCGGGCCGCCCAATCGACCTGCCGCGAGGAATCACT 89377

QY 124 ACCCAGCAATTTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 183

Db 89378 ACCCAGCAATTTGGGCTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 89437

QY 184 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACCTTGGACTTGGCG 243

Db 89438 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACCTTGGACTTGGCG 89497



QY 244 TCAGTGGACACCGAAACGTCGATCCGATATTGTGTATCAGCCACAGCGGGTTTTCAAAAGTG 303  
|||||  
Db 89498 TCAGTGGACACCGAAACGTCGATCCGATATTGTGTATCAGCCACAGCGGGTTTTCAAAAGTG 89557  
|||||  
QY 304 CGCCCACTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG 363  
|||||  
Db 89558 CGCCCACTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG 89617  
|||||  
QY 364 AATTTACGCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACCACAGTCCGATTG 423  
|||||  
Db 89618 AATTTACGCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACCACAGTCCGATTG 89677  
|||||  
QY 424 TGGGATCTTAACACAGACACACCGGACTTCACCTGCACAGGTCATAAGCAGTGGTTCCTG 483  
|||||  
Db 89678 TGGGATCTTAACACAGACACACCGGACTTCACCTGCACAGGTCATAAGCAGTGGTTCCTG 89737  
|||||  
QY 484 TCGGTATCTCTGGCTCGGATGGCAACGGTTGGCCAGCGGTTGCCAAAGCGGGCTCTATA 543  
|||||  
Db 89738 TCGGTATCTCTGGCTCGGATGGCAACGGTTGGCCAGCGGTTGCCAAAGCGGGCTCTATA 89797  
|||||  
QY 544 ATCATCTGGGACCCGGAGACGGGTCCAGAGAGGGGGCGACCCCTTGAGTGGGCACAGAAA 603  
|||||  
Db 89798 ATCATCTGGGACCCGGAGACGGGTCCAGAGAGGGGGCGACCCCTTGAGTGGGCACAGAAA 89857  
|||||  
QY 604 CACATCAACTGCCTCGCTCGGTAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGCT 663  
|||||  
Db 89858 CACATCAACTGCCTCGCTCGGTAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGCT 89917  
|||||  
QY 664 TCCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAAAAATTGGCCAGTGCCTTATG 723  
|||||  
Db 89918 TCCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAAAAATTGGCCAGTGCCTTATG 89977  
|||||  
QY 724 AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGCGGGCCTTATT 783  
|||||  
Db 89978 AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGCGGGCCTTATT 90037  
|||||  
QY 784 TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTGTGC 843  
|||||  
Db 90038 TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTGTGC 90097  
|||||  
QY 844 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATTCGCTGAGCACCGGATACGTC 903  
|||||  
Db 90098 CGGACGTTCTCTGGCCACGCTCACTGGGTAAACAACATTCGCTGAGCACCGGATACGTC 90157  
|||||  
QY 904 CTGCGCACTGGTCCATTCCATCCGTTGAAGGATCGCTCCAAAGAGCCACCTCAGTTTGAGC 963  
|||||  
Db 90158 CTGCGCACTGGTCCATTCCATCCGTTGAAGGATCGCTCCAAAGAGCCACCTCAGTTTGAGC 90217  
|||||  
QY 964 A-----C 965  
|  
Db 90218 AGTAAGAATTCATATGCTCTCTATATATGCTGTATATCTCTATAATCTCTATAAGAGATTGTTTACAGC 90277  
|||||  
QY 966 TCAGGAATTCAGGAATCTGCTTGAAGCGCTACCAGGCGGTGTCCTGACGAGTGGA 1025  
|||||  
Db 90278 TCAGGAATTCAGGAATCTGCTTGAAGCGCTACCAGGCGGTGTCCTGACGAGTGGA 90337  
|||||  
QY 1026 GTCGCTGGTTTCCTGTTCCGATGACAACACCCCTCTATCTGTGGCGGAACACCAGAACAA 1085  
|||||  
Db 90338 GTCGCTGGTTTCCTGTTCCGATGACAACACCCCTCTATCTGTGGCGGAACACCAGAACAA 90397  
|||||  
QY 1086 GTCGCTTGAGCGCATGACAGGCGCACAGACGTTGGTCAACGATGTGAATATTCGCCGGA 1145  
|||||  
Db 90398 GTCGCTTGAGCGCATGACAGGCGCACAGACGTTGGTCAACGATGTGAATATTCGCCGGA 90457  
|||||  
QY 1146 TGFAAAGCTAATTCGCTGCTGTTTCATTGACAAGTCAAGTGTGTCGCGAGCCAGCGGA 1205  
|||||  
Db 90458 TGFAAAGCTAATTCGCTGCTGTTTCATTGACAAGTCAAGTGTGTCGCGAGCCAGCGGA 90517  
|||||  
QY 1206 TGGTCAGTACATGGCCACCTTCGGGGTCAATGTGCAGGCTGTTTACACGGTGCCTGGTGC 1265  
|||||  
Db 90518 TGGTCAGTACATGGCCACCTTCGGGGTCAATGTGCAGGCTGTTTACACGGTGCCTGGTGC 90577  
|||||  
QY 1266 CGGGACTCCCGCTTGATTGTTTCCGGCAGCAAGACTCAACTCTAAA----- 1313

Db 90578 CGCGGACTCCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAAAGGTTAGCACGT 90637  
|||||  
QY 1314 -----AGTATGGAGTGTGCAGA 1330  
|||||  
Db 90638 AGTTTCAGTTTCTAGCGCTAGCTAATACATATCCCTCTGCTTAGTATGGAGTGTGCAGA 90697  
|||||  
QY 1331 CGAAGAAACTGGCACAGGAGCTGCCCTGGACATGGCGATGAGGTGTTCCGGAGTGGACTGGG 1390  
|||||  
Db 90698 CGAAGAAACTGGCACAGGAGCTGCCCTGGACATGGCGATGAGGTGTTCCGGAGTGGACTGGG 90757  
|||||  
QY 1391 CGCCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGACAAAGTTATAAAGCT----- 1442  
|||||  
Db 90758 CGCCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGACAAAGTTATAAAGCTGTAAGTTT 90817  
|||||  
QY 1443 -----ATGGGCTTA 1451  
|||||  
Db 90818 GAGTAACGAGGTTTCTATATAACAATATACAGACTCTAATTGTATCTTTCAGATGGGCTTA 90877  
|||||  
QY 1452 TTAACAAATCATTAACCTGTACACGCTAAGAAATACTTAGGAATAAAACGTCCT 1511  
|||||  
Db 90878 TTAACAAATCATTAACCTGTACACGCTAAGAAATACTTAGGAATAAAACGTCCT 90937  
|||||  
QY 1512 GAGTAAA 1518  
|||||  
Db 90938 GAGTAAA 90944  
|||||

RESULT 7  
AE003589/c 302473 bp DNA linear INV 04-OCT-2000  
LOCUS Drosophila melanogaster genomic scaffold 142000013386046 section 2  
DEFINITION of 16, complete sequence.  
ACCESSION AE003589 AE002638  
VERSION AE003589.2 GI:10727410  
KEYWORDS HTG.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 302473)  
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,  
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,  
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,  
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,  
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,  
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor  
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,  
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,  
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,  
Benos,P.V., Berman,B.P., Bhandari,D., Bolintineanu,D., Borkov,D.,  
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,  
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,  
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de  
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,  
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,  
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrier,S.,  
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,  
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,  
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,  
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,  
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,  
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,  
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,  
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,  
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,  
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,  
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,  
Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,  
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,  
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,  
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,



Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,  
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,  
Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,  
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,  
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
The genome sequence of Drosophila melanogaster  
Science 287 (5461), 2185-2195 (2000)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers  
1. 302473  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/Chromosome="2L"  
/gene="Plc21C"  
<3618. .>39613  
/note="CG4574"  
/map="21B7-21C1"  
/db\_xref="FLYBASE:FBan0004574"  
/db\_xref="FLYBASE:FBgn0004611"

gene

join(3618. .3958,11098. .11376,11383. .11475,19033. .19179,  
27475. .27612,28770. .28903,29373. .29558,29704. .29901,  
30371. .30553,31293. .31473,32562. .32796,34065. .34210,  
35933. .36146,36220. .37083,37396. .37843,37909. .38301,  
38363. .38507,38569. .39613)  
/gene="Plc21C"  
/product="CT14752"  
/db\_xref="FLYBASE:FBan0004574"  
/db\_xref="FLYBASE:FBgn0004611"  
join(11383. .11475,19033. .19179,27475. .27612,28770. .28903,  
29373. .29558,29704. .29901,30371. .30553,31293. .31473,  
32562. .32796,34065. .34210,35933. .36146,36220. .37083,  
37396. .37843,37909. .38301,38363. .38507,38569. .38802)  
/gene="Plc21C"  
/note="Plc21C gene product"  
/codon\_start=1  
/protein\_id="AAF51507.1"  
/db\_xref="GI:7296215"  
/db\_xref="FLYBASE:FBan0004574"  
/db\_xref="FLYBASE:FBgn0004611"  
/translation="MMSAGGTIISTASVEVPQALQDGEKFIKRWDDSGTGTPVTRMVD  
AKGFFLYVDQNNELDILDIAIRIVRTGQYAKRPDKLRIQIVTLGPQDTLEEKTVT  
VCHGSDFVNMTFVFCCTRRDIAQLMTDGLIKLAYSIAQLNGSAIMFLQRAHTKLCLO  
VDKSGRIPVKNIIKLFQNKEDKRKVEKALDVTGLPSGKVDSISVSKFEQFDFYNLYK  
YLTQRSEVERLFDISIVGNSKRKCMSIAQLVEFLNKTQRDLRLNEILYPYANPARAKEL  
IQQYEPNKNQAKGQLSLDGFRLYLMGDDNPIMAPSKLDLCDDMQPMHYFINSSHN  
TYLTGHQLTGKSSVEIYKQCLLAGRCVELDFWNGRTEEPVIVHGTYTFVPEIIPAKDVL  
EAIASAFKTSYPIVILSFENHCNPROQAKIANYCREIFGDMLLDKPLDSHPLEPNMD  
LPPPAMLRKIIKNNKKHHHHHHKKKPAQVGTTPAANNKLTANSVDAAKAAQVVG  
LSASHEDGGVTRSTANGDVATGTGSAAGTAGHAPLQOIROSCKDSTGSSSDSSS  
EDESLPNTTPNLPSGNEPPEKAQKETEAGAEISALVNVVQPIHFSFENAEKNRCY  
EMSSFDEKQATLLKERPIEFVYNNKQLSRVYPAGTRFDSNFMPLQFWNAGCQLVA  
LNFQTLDLAQMLNLGIFEYNARSGYLLKPEFMRRSDRRDLDPFAESTVDGIIAGTVSIT  
VLSGGFLTDRKRANTFVEVDMYGLPADTVRKFKRTKTVRDNGMNPPLYDEEFPFKKWL  
PELASIRIAAYEEGKLIGHRLVPVIGLCPGYRHNLRSEVGQPIALASLFLCVVVKD  
YVPDDLNSFAEALANIKYQSELEKRDIOQLSVLTDEAEALGSADEDSLSKCSGQKKELR  
PVESLTSPKHRPSISAAAMSDVTVDRTDGGRGDSISIVAPSIQHQSILDOQSVST  
SIRQVSSQFDVLVLAEPLEKILDHKSVEKREKLEKKLES LRKKHDKKIKIAGQK  
SSPLEGKKPKFAITNKLVRKLSNKSLSCLSPHSEPCVEIIPACPLDLGDSSESAADA  
GEDLAGGSSSLDQRTQESRKLRSACRETSQYRELQKYHEAIIYSAAEKVLKTSQTGQT  
KQLKAGSKVTGEVMHQLEARRNEVKNLATVHRDRDELIRMKREVAASSVVERGVAER  
VRLKQTFDRRTDELQKHDSVRNALAEHRSKARQILDKEAESRSCVSSNGFLVLFHGP  
HHHGCTGSGSALSNGNLTNLNDAGAGGSHSAISPAKSHNSIAAAEMKT"  
complement(<11712. .>12601)  
/gene="CG11912"

gene

complement(<11712. .>12601)  
/gene="CG11912"  
/map="21B7-21B8"  
/db\_xref="FLYBASE:FBan0011912"  
/db\_xref="FLYBASE:FBgn0031248"  
complement(11712. .12601)  
/gene="CG11912"  
/product="CT37080"  
/db\_xref="FLYBASE:FBan0011912"  
/db\_xref="FLYBASE:FBgn0031248"  
complement(11773. .12588)  
/gene="CG11912"  
/note="CG11912 gene product"  
/codon\_start=1  
/protein\_id="AAF51509.1"  
/db\_xref="GI:7296217"  
/db\_xref="FLYBASE:FBan0011912"  
/db\_xref="FLYBASE:FBgn0031248"  
/translation="MKQFAVIFALALASVSALSVPQPGFPEGRIINGYEAAKGEAPYI  
VSLQTSNSHFCAGSLLDDEVITVTAHCLITYNQQAVAGAHSRDQENVQIRKFTNAQ  
YVIRHENYGGVGPNDIGLILKKEEDAFDLNAVARDGNSPVSASVSLPSKTFQGTSDGYL  
YGWGRDNSGLLPLNLQKLDAIIVDYNCKAALPSNNSLAETNVCTHTPGKADGSCNGD  
SGGPLYSQSSSRGAELIGIVSWGYPCLSTTYPVSVTSSSFLPWIDENRKA"

gene

complement(<13648. .>14580)  
/gene="CG11911"  
/map="21B7-21B8"  
/db\_xref="FLYBASE:FBan0011911"  
/db\_xref="FLYBASE:FBgn0031249"  
complement(13648. .14580)  
/gene="CG11911"  
/product="CT37076"  
/db\_xref="FLYBASE:FBan0011911"  
/db\_xref="FLYBASE:FBgn0031249"  
complement(13743. .14576)  
/gene="CG11911"  
/note="CG11911 gene product"  
/codon\_start=1  
/protein\_id="AAF51508.1"  
/db\_xref="GI:7296216"  
/db\_xref="FLYBASE:FBan0011911"  
/db\_xref="FLYBASE:FBgn0031249"  
/translation="MKLITVTLVIALVAAAQAKLSDKLAKLVPSFATGFVINGTEAE  
PHSAPYIVSLATNYLKHSHICGGTFLINKDWITAAHCISEPVGMSIIAGLHTRAEVDE  
LTQQQVDFGRVHEKTYGGVGPXDIALHVNESFIFNEWVOPATLPSREQVHEGETHL  
YGWGPKSYIFSGAKTLQTVTTOILNVECKELPESAPIAESLSSSLQSQSKSACN  
GDSGGPLVVEFTNAPSELIGIVSWGYPICGLANWPSIYTKVSAIDWITNIQSAYYKL  
F"  
complement(<49281. .>51230)  
/gene="CG11907"  
/map="21C1-21C1"  
/db\_xref="FLYBASE:FBan0011907"  
/db\_xref="FLYBASE:FBgn0031250"  
complement(join(49281. .49459,49730. .50200,50254. .50554,  
50786. .>51230))  
/gene="CG11907"  
/product="CT33136"  
/db\_xref="FLYBASE:FBan0011907"  
/db\_xref="FLYBASE:FBgn0031250"  
complement(join(49281. .49459,49730. .50200,50254. .50554,  
50786. .51049))  
/gene="CG11907"  
/note="CG11907 gene product"  
/codon\_start=1  
/protein\_id="AAF51506.1"  
/db\_xref="GI:7296214"  
/db\_xref="FLYBASE:FBan0011907"  
/db\_xref="FLYBASE:FBgn0031250"  
/translation="MPGTSDDRLLVDALDEDEILGTEGSFVPAFRPQDHHADEQCLM  
ERQQDEVLVSNESGRLFTYLVFYLIGTMTPNFFVTAEDYWKYKFNASINNT  
DLEELTPLQKSTCDLALTATISGTTFLLLNAIFGHVSLRUKMLGTLMWLILFGV  
TTGFEINTDTWQEOFFLITLIIVVLLNISAATWSGALYVAGLFPSEFTAVWSGOA  
LGGILTALAFILVLAFTDGPNTTAFIFFIVGGVLILLCTVCYVILARKPFFRYLEGG  
DKYKVIKTRAVPSHNRNGSAEGLPLEPILRQVMSKIYHLAISLALLYTTILTSVPAVTVL  
MQSEYCHSVWTAMVMYFALSNGYFTNILLINAPKRVKQHEKELASSIMAAALSCGMAY  
GSLLSLVFVQML"

gene

complement(<11712. .>12601)  
/gene="CG11912"

complement(<11712. .>12601)  
/gene="CG11912"  
/product="CT37080"  
/db\_xref="FLYBASE:FBan0011912"  
/db\_xref="FLYBASE:FBgn0031248"  
complement(11773. .12588)  
/gene="CG11912"  
/note="CG11912 gene product"  
/codon\_start=1  
/protein\_id="AAF51509.1"  
/db\_xref="GI:7296217"  
/db\_xref="FLYBASE:FBan0011912"  
/db\_xref="FLYBASE:FBgn0031248"  
/translation="MKQFAVIFALALASVSALSVPQPGFPEGRIINGYEAAKGEAPYI  
VSLQTSNSHFCAGSLLDDEVITVTAHCLITYNQQAVAGAHSRDQENVQIRKFTNAQ  
YVIRHENYGGVGPNDIGLILKKEEDAFDLNAVARDGNSPVSASVSLPSKTFQGTSDGYL  
YGWGRDNSGLLPLNLQKLDAIIVDYNCKAALPSNNSLAETNVCTHTPGKADGSCNGD  
SGGPLYSQSSSRGAELIGIVSWGYPCLSTTYPVSVTSSSFLPWIDENRKA"

complement(<13648. .>14580)  
/gene="CG11911"  
/map="21B7-21B8"  
/db\_xref="FLYBASE:FBan0011911"  
/db\_xref="FLYBASE:FBgn0031249"  
complement(13648. .14580)  
/gene="CG11911"  
/product="CT37076"  
/db\_xref="FLYBASE:FBan0011911"  
/db\_xref="FLYBASE:FBgn0031249"  
complement(13743. .14576)  
/gene="CG11911"  
/note="CG11911 gene product"  
/codon\_start=1  
/protein\_id="AAF51508.1"  
/db\_xref="GI:7296216"  
/db\_xref="FLYBASE:FBan0011911"  
/db\_xref="FLYBASE:FBgn0031249"  
/translation="MKLITVTLVIALVAAAQAKLSDKLAKLVPSFATGFVINGTEAE  
PHSAPYIVSLATNYLKHSHICGGTFLINKDWITAAHCISEPVGMSIIAGLHTRAEVDE  
LTQQQVDFGRVHEKTYGGVGPXDIALHVNESFIFNEWVOPATLPSREQVHEGETHL  
YGWGPKSYIFSGAKTLQTVTTOILNVECKELPESAPIAESLSSSLQSQSKSACN  
GDSGGPLVVEFTNAPSELIGIVSWGYPICGLANWPSIYTKVSAIDWITNIQSAYYKL  
F"  
complement(<49281. .>51230)  
/gene="CG11907"  
/map="21C1-21C1"  
/db\_xref="FLYBASE:FBan0011907"  
/db\_xref="FLYBASE:FBgn0031250"  
complement(join(49281. .49459,49730. .50200,50254. .50554,  
50786. .>51230))  
/gene="CG11907"  
/product="CT33136"  
/db\_xref="FLYBASE:FBan0011907"  
/db\_xref="FLYBASE:FBgn0031250"  
complement(join(49281. .49459,49730. .50200,50254. .50554,  
50786. .51049))  
/gene="CG11907"  
/note="CG11907 gene product"  
/codon\_start=1  
/protein\_id="AAF51506.1"  
/db\_xref="GI:7296214"  
/db\_xref="FLYBASE:FBan0011907"  
/db\_xref="FLYBASE:FBgn0031250"  
/translation="MPGTSDDRLLVDALDEDEILGTEGSFVPAFRPQDHHADEQCLM  
ERQQDEVLVSNESGRLFTYLVFYLIGTMTPNFFVTAEDYWKYKFNASINNT  
DLEELTPLQKSTCDLALTATISGTTFLLLNAIFGHVSLRUKMLGTLMWLILFGV  
TTGFEINTDTWQEOFFLITLIIVVLLNISAATWSGALYVAGLFPSEFTAVWSGOA  
LGGILTALAFILVLAFTDGPNTTAFIFFIVGGVLILLCTVCYVILARKPFFRYLEGG  
DKYKVIKTRAVPSHNRNGSAEGLPLEPILRQVMSKIYHLAISLALLYTTILTSVPAVTVL  
MQSEYCHSVWTAMVMYFALSNGYFTNILLINAPKRVKQHEKELASSIMAAALSCGMAY  
GSLLSLVFVQML"

gene	<68682..>77995 /gene="al" /note="CG3935" /map="21C1-21C2" /db_xref="FLYBASE:FBan0003935" /db_xref="FLYBASE:FBgn0000061" join(68682..69037,75067..75450,76257..76302,76864..77132,77259..77995) /gene="al" /product="CT12963" /note="Nucleotide sequence of the Celera sequence differs			
	Query Match	82.7%;	Score 1285.8;	DB 3; Length 302473;
	Best Local Similarity	89.2%;	Pred. No. 6.6e-311;	
	Matches 1505;	Conservative 0;	Mismatches 7;	Indels 175; Gaps 4;
mRNA	7	CAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCGC	66	
	258606	CAGAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCGC	258547	
	67	CTCGTTT---ACACGGGGGAGGAAGCCGGCCCCGCCAATCGACCTGCCGGCAGGAATCACT	123	
	258546	CTCGTTTCGGACACCGGGCAGGAAGCGGGCCCCGCCAATCGACCTGCCGGCAGGAATCACT	258487	
	124	ACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT	183	
	258486	ACCCAGCAATTGGGTCTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT	258427	
	184	TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTTGGACTTGGCG	243	
	258426	TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTTGGACTTGGCG	258367	
	244	TCAGTGGACACCGAAAACGTGATCGATATTTGTATATCAGCCACAGCGCGTTTCAAAGTG	303	
	258366	TCAGTGGACACCGAAAACGTGATCGATATTTGTATATCAGCCACAGCGCGTTTCAAAGTG	258307	
	304	CGCCCACTGACAAGATGCACGAGTTCCATCGCCGAGTGGCAAGTGGCCACACCACTGCGATTG	423	
	258306	CGCCCACTGACAAGATGCACGAGTTCCATCGCCGAGTGGCAAGTGGCCACACCACTGCGATTG	258247	
	364	AATTTCAGCCCCGGATGGTCTCATCTCGCCAGTGGAAAGTGGCCACACCACTGCGATTG	483	
	258246	AATTTCAGCCCCGGATGGTCTCATCTCGCCAGTGGAAAGTGGCCACACCACTGCGATTG	258187	
	424	TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTGATATAAGCAGTGGTTCTG	483	
	258186	TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTGATATAAGCAGTGGTTCTG	258127	
	484	TGCGTATCCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAGCGGCTCTATA	543	
	258126	TGCGTATCCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAGCGGCTCTATA	258067	
	544	ATCATCTGGGACCCGGAGACGGGTACGACGAAGAAGGGCGACCCCTGAGTGGGCACAAGAAA	603	
	258066	ATCATCTGGGACCCGGAGACGGGTACGACGAAGAAGGGCGACCCCTGAGTGGGCACAAGAAA	258007	
	604	CACATCAACTGCCCTCGCCTGGGAACCGTATCATTCGGATCCGGAGTGCAGGAAACTTGCT	663	
	258006	CACATCAACTGCCCTCGCCTGGGAACCGTATCATTCGGATCCGGAGTGCAGGAAACTTGCT	257947	
	664	TCCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAAAATTGGGCCAGTGCCTTATG	723	
	257946	TCCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAAAATTGGGCCAGTGCCTTATG	257887	
	724	AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGGTGGAGCGGCGCTTATT	783	
	257886	AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGGTGGAGCGGCGCTTATT	257827	
	784	TATACATCCTCCAAAGATCCGACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC	843	
	257826	TATACATCCTCCAAAGATCCGACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC	257767	
	844	CGGACGTTCTCTGGCCCAAGCTCACTGGGTAAACAACATTGGCGTGAGCACCGATTACGTC	903	

Db	257766	CGGACGTTCTCTGGCCACGCTCACTACCTGGGTAACAACAACATTGCGCTGAGCACCGACTAGTCT	257707
QY	904	CTGCGCACTGGTCCCATTCATCCGGTGAAGGATCGCTCCAAGAGACCCACCTCAGTTTGAGC	963
Db	257706	CTGCGCACTGGTCCCATTCATCCGGTGAAGGATCGCTCCAAGAGACCCACCTCAGTTTGAGC	257647
QY	964	A-----C 965	
Db	257646	AGTAAGAATTCATATGCTCTCTATATATATGCTGTATATCTCTAAAGAGATGTTTACAGC	257587
QY	966	TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTGCCCTGACGAGGTGGA	1025
Db	257586	TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTGCCCTGACGAGGTGGA	257527
QY	1026	GTCGCTGGTTTCCTGTTCGGATGACAACAACCCCTCTATCTGTGGCGGAACAACCCAGAACAA	1085
Db	257526	GTCGCTGGTTTCCTGTTCGGATGACAACAACCCCTCTATCTGTGGCGGAACAACCCAGAACAA	257467
QY	1086	GTCGCTTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGATGTGAATATTCGCCGGA	1145
Db	257466	GTCGCTTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGACGTGAATATTCGCCGGA	257407
QY	1146	TGTAAAGCTAATTGCGTCTGCTTCATTGACAAAGTCAGTGCCTGTGTGGCGAGCCAGCGGA	1205
Db	257406	TGTAAAGCTAATTGCGTCTGCTTCATTGACAAAGTCAGTGCCTGTGTGGCGAGCCAGCGGA	257347
QY	1206	TGTCAGTACATGGCCACCTTCCGGGGTCATGTGCAGGCTGTTTACACGGTTCCTTGGTC	1265
Db	257346	TGTCAGTACATGGCCACCTTCCGGGGTCATGTGCAGGCTGTTTACACGGTTCCTTGGTC	257287
QY	1266	CGCGGACTCCCGCTTGATTGTTTCCGGGACGAAAGACTCAACTCTAAA-----1313	
Db	257286	CGCGGACTCCCGCTTGATTGTTTCCGGGACGAAAGACTCAACTCTAAAAGGTTAGCACGT	257227
QY	1314	-----AGTATGGAGTGTGCAGA 1330	
Db	257226	AGTTTCAGTTTCTTAGCGCTAGCTAATACATATCCTCTTGTCTAGTATGGAGTGTGCAGA	257167
QY	1331	CGAAGAAACTGGCACAGGAGTGCCTGGACATCGGGATGAGGTGTTCCGAGTGGACTGGG	1390
Db	257166	CGAAGAAACTGGCACAGGAGTGCCTGGACATCGGGATGAGGTGTTCCGAGTGGACTGGG	257107
QY	1391	CGCCCGATGGCTTAGAGTTGCCTCTGTGGCGAAGGACAAAAGTTATAAAGCT-----1442	
Db	257106	CGCCCGATGGCTTAGAGTTGCCTCTGTGGCGAAGGACAAAAGTTATAAAGCTGTAAAGTT	257047
QY	1443	-----ATGGGCTTA 1451	
Db	257046	GAGTAACGAGGTTTCTATAACAATATACAGACTCTAATTGTATCTTTCAGATGGGCTTA	256987
QY	1452	TTAACAAATCATTAACCTGTACACGGTAAGAAAATACTTAGGAATAAAGTAAACGTCCT	1511
Db	256986	TTAACAAATCATTAACCTGTACACGGTAAGAAAATACTTAGGAATAAAGTAAACGTCCT	256927
QY	1512	GAGTAAA 1518	
Db	256926	GAGTAAA 256920	

RESULT 8  
AX135811  
LOCUS AX135811 1859 bp DNA linear PAT 29-MAY-2001  
DEFINITION Sequence 5 from Patent WO0132614.  
ACCESSION AX135811  
VERSION AX135811.1 GI:14272046  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1859)  
AUTHORS Utku,N.

TITLE		Novel genes tzap7/a, tzap7/b and tzap7 involved in t cell activation and uses thereof									
JOURNAL		Patent: WO 0132614-A 5 10-MAY-2001; Utku, Nalan (DE)									
FEATURES		Location/Qualifiers									
source		1..1859									
		/organism="Homo sapiens"									
		/db_xref="taxon:9606"									
CDS		20..1477									
		/note="unnamed protein product"									
		/codon_start=1									
		/protein_id="CAC39576.1"									
		/db_xref="GI:14272047"									
		/translation="MAAAVADEAVARDVQRLLVQFQDEGGQLLGSFPDVPVDITPDRL QLVCNALLAQEDPLPLAFFVHDAEIVSSLGKLTLESQAVETEKVLIDIIYQPAIFRVRA VTRCTSSLEGHSEAVISAFPTGKYLASGSDTTVRFWDLSTETPHFTCKGHRHWVL SISWSPDGKKLASGCKNGQIILLWDPSTGKQVGRTLAGHSKWITGLSWEPLHANPECRY VASSKDGSVRIWDTTAGRCERILTGHTQSVTCLRWGGDGLLYSASQDRTIKVWRAHD GVLCRTLQGHGHWNTMALSTDYALRTGAFEPAEASVNPQDLQGSLOELKERALSRYN LVRGQPERLVSGSDDFTLFLWSPAEDKKPLTRMTGHQALINQVXFSPDSRIVASASE DKSILWDGRTGKYLASLRGHVAAYQIAWSADSRLLVSGSSDSTLKVWDVKAQKLAM DLPGHAEVYAVDWSPDQQRVASGGKDKLRWRR"									
BASE COUNT		392 a	545 c	566 g	355 t	1 others					
ORIGIN											
		Query Match 30.7%; Score 478; DB 6; Length 1859;									
		Best Local Similarity 59.8%; Pred. No. 8.5e-109;									
		Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;									
QY		54	GATACAGGCGCCCTCGTTTACACGGCGGAGGAAGCCGCGCCGCAATCGACCTGCCGCG	113							
Db		70	GCTAGTGCAGTTCAGGATGAGGCGGCGCAGCTGCTGGGTTCCTCCGTTCCGACGTGCCCGT	129							
QY		114	AGGAATCACTACCCAGCAATTGGGACTGATTGCAACGCGCTGCTGAAAAACGAGGAAGC	173							
Db		130	GGACATCACCCCGGACAGGCTCGTGTGCAACGCGCTACTGGCCAGGAGGATCC	189							
QY		174	CACCTCATATTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTT	233							
Db		190	CCTGCCACTGGCTTCTTTGTCCACGATGCTGAGATCGTCTCCTCACTGGGGAAGACGTT	249							
QY		234	GGACTTGGCGTCAGTGGACACCGAAAAACGTGATCGATATTGTGTATCAGCCACAGGCGST	293							
Db		250	GGAGTCCCAGGCAAGTGAGACAGAGAAGGTCCTAGACATCATCTACCAGCCACAGGCTAT	309							
QY		294	TTTCAAAGTGCGCCCACTGCACAAAGATGCACGAGTTCCTATGCCGGGACACGCCGAGGCTGT	353							
Db		310	CTTCAGAGTCCGGGCTGTGACTCGCTGCACAGCTCCTTGAGGGTTCACAGTGAGGCAGT	369							
QY		354	GGTTTCGCTGAATTCAGCCCGGATGGTGTCTCATCTCGCCAGTGGAAAGTGGCGGACACCCAC	413							
Db		370	CATTTCGTGGCCTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACCCAC	429							
QY		414	AGTCCGATTGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTTCATAAGCA	473							
Db		430	CGTCCGCTTCTGGGATCTCAGCACAGACACACACATTTACATGCAAGGCACACAGACA	489							
QY		474	GTGGGTTCTGTGCGTATCCTGGGCTCCGGATGGCAACCGGTTGGCCAGCGGTTGCCAAGC	533							
Db		490	CTGGGTCCTTAGTATATCTGGTCTCCAGATGGCAAGAAGCTGGCCTCAGGCTGCAAGAA	549							
QY		534	GGGCTCTATAATCATCTGGGACCCGGAGACGGGTTCAGCAAGAGGGCGGACCCCTTGAGTGG	593							
Db		550	TGGCCAGATTCTCCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGGCGAGGACCCCTCGCTGG	609							
QY		594	GCACAAGAAACACATCAACTGCCCTGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAG	653							
Db		610	CCACAGCAAGTGGATCACAGGCCCTGAGCTGGGAGCCCTCCATCGGAACCCCTGAGTGCCG	669							
QY		654	GAAACTTGCTTCCGCCAGTGGAGACGGGGACTGCCGGATTTGGGACGTAAATTTGGGCCA	713							
Db		670	CTATGTGGCCAGCAGCTCCAAGGATGGCAGTGTGCGGATCTGGGACACAACTGCAGGCCG	729							

QY	714	GTGCCTTATGAACATTGCCGGACACACAAATCTGTGACACAGTGAAGTGGGTGGAGC	773
Db	730	CTGTGAGCGCATCCTCACCGGGACACACCCAGTGGTCACTCTCTCCGGTGGGAGGGA	789
QY	774	GGGCTTATTTATACATCCTCCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG	833
Db	790	CGGCTTCTCTACTCTGCCCTCCAGGACCGCACCATCAAAGTCTGGAGAGCTCATGACGG	849
QY	834	AATCTTGTGCCGACGTTCTCTGGCCAAAGCTCACTGGGTAACAACAACTTGCCTGAGCAC	893
Db	850	TGTGCTGTGCCGACTCTGCAAGGCCACGCGCACTGGGTGAACACCATGGCCCTCAGCAC	909
QY	894	CGATTACGTCCTCGGCACACTGGTCCATTCCATCCGGTGAAGATCGCTCCAAGAGCC--A	950
Db	910	TGACTATGCCCTCGGCACACTGGGCGCTTTGAACCTGCTGAGGCTCAGTTAATCCCCAAGA	969
QY	951	CCTCAGTTTGAGCACTGAGGAATGCAAGAAATGTCCTTGAAGCGCTACAGGCCGTGTG	1010
Db	970	CCTCAAAGGATCCTTGCAAGGAGTTGAAGGAGAGGGCTCTGAGCGGATACAACTCGTGG	1029
QY	1011	CCCTGACGAGGTGGAGTCGCTGGTTTCTGTTCCGATGACAAACACCCCTCTAUCTGTGG--	1068
Db	1030	GGCCAGGGTCCAGAGAGGCTGGTGTCTGGCTCGGACGACTTCACCTTATTCTCTGTGGTC	1089
QY	1069	-CGGAACAACACAGAAAGTGGCTTGAGCGCATGACAGGGCACAGAACTGGTCAACGA	1127
Db	1090	CCACGACAGAGGACAAAAGCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA	1149
QY	1128	TGTGAAATATTCCGCGGATGTAAGCTAATTGGCTCTGCTTCATTGACAAAGTCAGTGG	1187
Db	1150	GGTGNCTTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCCTTTGACAAAGTCCATCAA	1209
QY	1188	TCTGTGGCGAGCCAGCGATGGTCAGTACATGGCCACCTTCCGGSGTTCATGTGCAGGCTGT	1247
Db	1210	GCTGTGGGATGGCAGACGGCAAGTACCTGGCTTCCCTACGCGGCCACGTGGTGGCGGT	1269
QY	1248	TTACACGGTTGCCGTGGTCCGCGGACTCCCGCTTGATTTTCCGGCAGCAAAAGACTCAAC	1307
Db	1270	GTACAGATTGCGTGGTGCAGCTGCAGTCCGCTCGGTCAGCGGCAGCAGTGACAGCAC	1329
QY	1308	TCTAAAGATATGGAGTGTGCAGACGAAGAAACTGGCACAGGAGTGCCTGGACATGCGGA	1367
Db	1330	ACTGAAGTGTGGGATGTGAAGGCCACAGAAGCTGGCCATGGACCTCCCGGCCAGCGGA	1389
QY	1368	TGAGGTGTTCCGAGTGGACTGGGCGCCCGATGGCTCTAGAGTTGCTTGGTGGCAAGGA	1427
Db	1390	TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA	1449
QY	1428	CAAAATTATAAAGCTATGG	1446
Db	1450	CAAAATGCCCTCCGATATGG	1468

RESULT 9	AK001320	1853 bp	mRNA	linear	PRI 01-AUG-2002
LOCUS	Homo sapiens cDNA FLJ10458 fis, clone NT2RP1001457, highly similar to Homo sapiens partial mRNA for beta-transducin family protein.				
DEFINITION	AK001320				
ACCESSION	AK001320.1 GI:7022501				
VERSION	oligo capping; fis (full insert sequence).				
KEYWORDS	Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,				
SOURCE	clone_lib:NT2RP1 clone:NT2RP1001457.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1				
AUTHORS	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,				







AX077627  
LOCUS AX077627 1854 bp DNA linear PAT 22-FEB-2001  
DEFINITION Sequence 82 from Patent WO0105970.  
ACCESSION AX077627  
VERSION AX077627.1 GI:13122009  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1854)  
AUTHORS Yue,H., Tang,Y.T., Bandman,O., Hillman,J.L., Lal,P., Au-Young,J.,  
Reddy,R., Yang,J., Baughn,M.R., Lu,D.A., Azimzai,Y. and  
Patterson,C.  
TITLE Gtp-binding protein associated factors  
JOURNAL Patent: WO 0105970-A 82 25-JAN-2001;  
Incyte Genomics, Inc. (US)  
FEATURES  
source Location/Qualifiers  
1..1854  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/note="Incyte ID No: 2216640CB1"  
BASE COUNT 389 a 548 c 563 g 354 t  
ORIGIN  
Query Match 30.7%; Score 477.4; DB 6; Length 1854;  
Best Local Similarity 59.8%; Pred. No.1.2e-108;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;  
QY 54 GATACAGGCGCGCTCGTTTACACGGGCGAGGAAGCGCCGCAATCGACCTGCCGGC 113  
DB 69 GCTAGTGCAGTTCCAGGATGAGGGCGGCAGCTGCTGGGTTCCCGCTTCGACGTGCCCGT 128  
QY 114 AGGAATCACTACCCAGCAATTGGGACTGATTGCAACGCGCTGCTGTAACAAACGAGGAAGC 173  
DB 129 GGACATCACCCCGACAGGCTGCAGCTCGTGTGCAACGCGCTACTGGCCCGAGGAGATCC 188  
QY 174 CACTCCATATTTGTTTTGCGTGGCGAGGATGAGATCAAGAAGACCGCTGGAGGACAGCTT 233  
DB 189 CCTGCCACTGGCTTTCTTTGTCCACGATGCTGAGATCGTCTCCTACTGGGGAAGACGTT 248  
QY 234 GGACTTGGCGTCAGTGGACACCGAAACGTCATCGATATTGTGTATCAGCCACAGGCGGT 293  
DB 249 GGAGTCCAGGCAGTGGAGACAGAGAAGGTCTCTAGACATCATCTACCAGCCACAGGCTAT 308  
QY 294 TTTCAAAGTCCGCCAGTGACAAGATGCACGAGTTCATCCGCGGACACCGCGAGGCTGT 353  
DB 309 CTTCAGAGTCCGGGCTGTGACTCGCTGCACACGCTCTTGAGGGTTCACAGTGAGGCAGT 368  
QY 354 GGTTCGCTGAATTTTCAGCCCCGGATGGTCTCATCTCGCCAGTGGAGTGGCGACACCAC 413  
DB 369 CATTCTGTGGCCTTCAGCCCTACGGGAAGTACCTGGCCAGTGGCTCTGGAGACACCAC 428  
QY 414 AGTGGATTGTGGGATCTTAACACAGAGACACCGGCACCTTCACCTGCACAGGTCATAAGCA 473  
DB 429 CGTGGCTTCTGGGATCTCAGACACAGACACACCACATTTCACATGCAAGGGACACAGACA 488  
QY 474 GTGGGTTCTGTGCGTATCCTGGGCTCCGATGGCAACCGTTGGCCAGCGGTTGCCAAAGC 533  
DB 489 CTGGGTCCTTAGTATATCTGTTCTCCAGATGGCAAGAAGCTGGCCCTCAGGCTGCAAGAA 548  
QY 534 GGGCTCTATAATCATCTGGGACCCCGAGACGGGTACGACAGAAGGGGCGACCCCTTGAGTGG 593  
DB 549 TGGCCAGATTCTCCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGGAGGACCCCTCGCTGG 608  
QY 594 GCACAAGAAACACATCAACTGCCTCGCTGGGAACCGTATCATCGCGATCCGGAGTGCAG 653  
DB 609 CCACAGCAAGTGGATCAGAGCCCTGAGCTGGGAGCCCTCCATCGGACCCCTGAGTGCCG 668  
QY 654 GAAACTTGCTCCGCCAGTGGAGACGGGACTGCCGGATTGGGACGTAAATTTGGGCCA 713  
DB 669 CTATGTGGCCAGCAGCTCCAAGGATGGCAGTGTGGGATCTGGGACACAACTGCAGGCCG 728

QY 714 GTGCCTTATGAACATTGCCGGACACACACAAATGCTGTGACAGCAGTGGGTGGAGC 773  
DB 729 CTGTGAGCGCATCTCACCGGGCACACCCAGTCGGTACCTGTCTCCGTTGGGAGGGGA 788  
QY 774 GGGCCTTATTTATACATCTCCAAAGATFCGCACAGTGAAGATGTGGCAGCAGCTGATGG 833  
DB 789 CGGGCTTCTCTACTCTGCTCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG 848  
QY 834 AATCTTGCCCGGAGCTTCTTGCCCAAGCTCACTGGGTAAACAACATTCGCTGACGAC 893  
DB 849 TGTGCTGTGCCGACTCTGCAAGGCCACGCGCCACTGGGTGAACACCATGGCCCTCAGCAC 908  
QY 894 CGATTACGTCCTGCGCAGCTGGTCCATTCCATCCGCTGGAAGGATCGTCCCAAGAGCC--A 950  
DB 909 TGACTATGCCCCTGCGCAGCTGGGGCCCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCCAAGA 968  
QY 951 CCTCAGTTTGAGCAGCTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCCGCTGTG 1010  
DB 969 CCTCCAAGGATCCTTGCAGGAGTTGAAGGAGAGGGCTCTGAGCCCGATACACCTCGTGGC 1028  
QY 1011 CCCTGACGAGGTGGAGTCGCTGGTTTCCCTGTTCCGATGACAAACACCCCTCTATCTGTGG-- 1068  
DB 1029 GGGCCAGGTTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACCTTATCTCTGTGGTC 1088  
QY 1069 -CGGAACAACACAGAACTGCGTTGAGCGCATGACAGGCGACCAAGAGTGGTCAACGA 1127  
DB 1089 CCCAGCAGAGGACAAAAGCCTCTCAGTCGGATGACAGGACACCAAGCTCTCATCAACCA 1148  
QY 1128 TGTGAAATATTGCCCGGATGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTGCG 1187  
DB 1149 GGTGCTCTTCTCTCCTGACTCCCGCATCGTGGCTAGTGCCTCCTTTGACAAGTCCATCAA 1208  
QY 1188 TCTGTGGCGAGCCAGCGATGGTCACTACATGAGCCACCTTCCGGGGTCAATGTGAGGCTGT 1247  
DB 1209 GCTGTGGATGGCAGGACGGCAAGTACCTGGCTTCCCTACGCGGCCAGCTGGTGGCTG 1268  
QY 1248 TTACACGGTTCGCTGGTCCGCGGACTCCCGCTTGATTGTTCCGGCAGCAAAAGACTCAAC 1307  
DB 1269 GTACCAGATTGCGTGGTCACTGACAGTCGCTCGCTCCTGTCAGCGGCAGAGTGACAGCAC 1328  
QY 1308 TCTAAAGTATGGAGTGTGCAGACGAAGAACTGGCACAGGAGCTGCCTGGACATGCCGA 1367  
DB 1329 ACTGAAGGTGTGGGATGTGAAGGCCCCAGAAAGCTGGCCATGGACCTGCCCGCCACGCCGA 1388  
QY 1368 TGAGGTGTTGCGAGTGGACTGGGCGCCCGATGGCTCTAGAGTTGCCCTGCTGGTGGCAAGGA 1427  
DB 1389 TGAGGTATATGCTGTTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA 1448  
QY 1428 CAAAGTTATAAGCTATGG 1446  
DB 1449 CAAATGCCCTCCGGATATGG 1467

RESULT 11  
BC002884  
LOCUS BC002884 1870 bp mRNA linear PRI 12-JUL-2001  
DEFINITION Homo sapiens, hypothetical protein similar to beta-transducin family, clone IMAGE:3943492, mRNA, partial cds.  
ACCESSION BC002884  
VERSION BC002884.1 GI:12804062  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1870)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

Contact: MGC help desk  
Email: cgaps-re@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,  
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 15 Row: 9 Column: 15  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7022501.

FEATURES  
source

Location/Qualifiers  
1. .1870  
/organism="Homo sapiens"  
/db\_xref="LocusID:54475"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3943492"  
/tissue\_type="Lung, small cell carcinoma"  
/clone\_lib="NIH\_MGC\_7"  
/lab\_host="DH10B-R"  
/note="Vector: pOT7"  
<1. .1465  
/codon\_start=2  
/product="hypothetical protein similar to beta-transducin  
family"  
/protein\_id="AAH02884.1"  
/db\_xref="GI:12804063"  
/translation="ARGAAVADEAVARDVORLLVQFQDEGGQLLGSPFDPVVDITPD  
RLQVNCNALLAQEDPLPAFFVHDAEIVSSLGKTLSEQAVETEKVIDIILYQQAIFRV  
RAVTRCTSSLEGHSEAVISVAFSPGKYLASGSDTTFVFWDLSTETPHEFTCKGHRHW  
VLSISWSPDGKKLASGCKNGQILLWDPSTGKQVGRFLAGHSKWITGLSWEPLHANPEC  
RYVASSKDGSVRIWDTAGRCERILTGHQSVTCRLRWGGDGLLYSASQDRTIKVWRA  
HDGVLCTRLOGHGHWNTMALSTDYALRTGAFEPAEASVNPQDLQGSLOELKERALS  
YNLVRGQPERLVSGSDFTFLWSPAEKDKPLFRMTGHQALINOVLFSPDSRVIVASA  
SFDKSIKLWDGRTGKYLASLRGHVAAYVQIAWSADSRLLVSGSSDSTLKVWDVKAQKL  
AMDLPGHAEVYAVDWSPDGQRVASGKDKCLRINRR"

BASE COUNT 415 a 541 c 562 g 352 t  
ORIGIN

Query Match		30.7%;	Score 477.4;	DB 9;	Length 1870;
Best Local Similarity		59.8%;	Pred. No. 1.2e-108;		
Matches 837;		Conservative 0;	Mismatches 556;	Indels 6;	Gaps 2;
QY	54	GATACAGGCGCGCTCGTTTACACGGGCGAGGAGCGCGCCGCAATCGACCTGCCGCG	113		
Db	58	GCTAGTCAGTTCAGGATGAGGGCGGGCAGCTGCTGGTTCCTCCGCTCGACGTGCCCGT	117		
QY	114	AGGATCACTACCCAGCAATTGGGACTGATTGCAACGCGCTGCTGAAACGAGGAAGC	173		
Db	118	GGACATCACCCCGACAGGCTGCAGCTCGTGTGCAACGCGCTACTGGCCCGAGGATCC	177		
QY	174	CACATCCATATTTGTTTTCGTTGGGCGGAGGATGAGATCAAGAGAGCCCTGGAGGACACGTT	233		
Db	178	CCTGCCACTGGCTTCTTTGTCCACGATGCTGAGATCGTCTCCTCACTGGGGAAGACGTT	237		
QY	234	GGACTTGGCGTCAGTGGACACCCGAAACCGTGATCGATATTGTGTATCAGCCACAGGCGGT	293		
Db	238	GGAGTCCCAGGCACTGGAGACAGAGAAGGTCTCTAGACATCATCTACCAGCCACAGGCTAT	297		
QY	294	TTTCAAAGTGCGCCAGTGACAGATGCACAGATTCCATGCCGGGACACCGCGAGGCTGT	353		

Db	298	CTTCAGAGTCCGGGCTGTGACTCGCTGCACCAGCTCCTTGGAGGGTCAAGTGAAGGCGAGT	357		
QY	354	GGTTTCGGCTGAATTTACGCCCGGATGGTGTCTCATCTCGCCAGTGGAAAGTGGCGACACAC	413		
Db	358	CATTTCTGTGGCCTTACGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACAC	417		
QY	414	AGTGCATTTGTGGGATCTTAACACACAGACACCCGACTTCACCTGCACAGGTCTATAAGCA	473		
Db	418	CGTGGCTTCTGGGATCTCAGCACACAGACACCAACATTTACATGCAAGGGACACAGACA	477		
QY	474	GTGGTTCGTGCGTATCTCTGGGCTCCGGATGGCAAACGGTTGGCCAGCGTTGCAAAGC	533		
Db	478	CTGGTTCCTTAGTATATCTCTGGTCTCCAGATGGCAAGAGCTGGCCTCAGGCTGCAAGAA	537		
QY	534	GGGCTCTAATAATCATCTGGGACCCGGAGACGGGTACGACAGAAGGGGCGACCCCTTGAGTGG	593		
Db	538	TGGCCAGATTCTCCTCTGGGACCCCAAGCACAGGAAACAGAGGTGGGACGACCTCGCTGG	597		
QY	594	GCACAAGAAACACATCAACTGCCCTGGCTGGGAACCGTATCATCGGATCCGGAGTGCAG	653		
Db	598	CCACAGCAAGTGGATCACAGGCCCTGAGTGGGAGCCCTCCATGCGAAACCTGAGTGCCG	657		
QY	654	GAACACTTCTCCGCCAGTGGAGACGGGGACTGCCGATTTGGGACGTAAATTTGGGCCA	713		
Db	658	CTATGTGCCAGCAGCTCCAAGGATGGCAGTGTGCGGATCTGGGACACAACTGCAGGCCG	717		
QY	714	GTGCCTTATGAACATTTGCCGGACACACAAATGCTGTACAGCAGTGAAGATGTGGGAGCAGCTGATGG	773		
Db	718	CTGTGAGCGCATCTCACCGGGCACACCCAGTCGGTGCACCTGTCTCCGGTGGGAGGGGA	777		
QY	774	GGCCTTATTTATACATCTCTCCAAAGATCGCACAGTGAAGATGTGGGAGCAGCTGATGG	833		
Db	778	CGGGCTTCTCTACTCTGCCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG	837		
QY	834	AATCTTGTGCCGAGCTTCTCTGGCCAAGCTCACTGGGTAAACAAATTCGCTGAGCAC	893		
Db	838	TGTGCTGTGCCGACTCTGCAAGGCCACGGCCACTGGGTGAACACCATGGCCCTCAGCAC	897		
QY	894	CGATTACGCTCCTGCGCACTGGTCCATTCCATCCGCTGAAGGATCGCTCCAAGAGCC--A	950		
Db	898	TGACTATGCTCCTGCGCACTGGGGCCTTTAAACCTGCTGAGGCCCTCAGTTAATCCCAAGA	957		
QY	951	CCTCAGTTTGAACACTCAGGAATTCGAGGAATCTGCCTTGAAGCGCTACCAGGCGGTGTG	1010		
Db	958	CCTCCAAGGATCCTTGCAGGAGTTGAAGGAGAGAGGGCTCTGAGCCGATACAACCTCGTGG	1017		
QY	1011	CCCTGACGAGGTGGAGTCGCTGTTTCTTCTGTTCCGATGACAACACCTCTATCTGTGG--	1068		
Db	1018	GGCCAGGGTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACTTATTCCTGTGGTC	1077		
QY	1069	-CGGAACAACCAGAAAGTGCCTTGAAGCGCATGACAGGGCACCAAGAACGTTGTCACACGA	1127		
Db	1078	CCCAGCAGAGGACAAAAGCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA	1137		
QY	1128	TGTGAATATTCCCGGATGTAAAGCTAATTGCGTCTGCTTCAATTGACAAGTCAAGTGG	1187		
Db	1138	GGTCTCTTCTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCCTTTGACAAGTCCATCAA	1197		
QY	1188	TCTGTGGCGAGCCAGCGATGGTCACTACATGGCCACCTTCCGGGGTCAATGTGAGGCTGT	1247		
Db	1198	GCTGTGGGATGGCAGGACGGGCAAGTACCTGGCTTCCCTACGGCGCCACGTTGGCTGCCGT	1257		
QY	1248	TTACACGGTTGCCCTGGTCCCGGACTCCCGCTTGAATGTTTCCGGCAGCAAGACTCAAC	1307		
Db	1258	GTACCAGATTGCGTGGTCACTGACAGTCGGCTCCTGGTCAGCGGACGACTGACAGCAC	1317		
QY	1308	TCTAAAAGTATGGAGTGTGCAGACCAAGAAACTGGCACAGGAGTGCCTGGACATGCCGA	1367		
Db	1318	ACTGAAGGTGTGGGATGTGAAGGCCCAAGAGCTGGCCATGGACCTGCCCGCCACGCGGA	1377		
QY	1368	TGAGTGTTCGGAGTGGACTGSGGCGCCCGATGGCTCTAGAGTTGCCTCTGTTGGCAAGGA	1427		
Db	1378	TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGTGGCAAGTGGTGGGAAGGA	1437		

QY 1428 CAAAGTTATAAAGCTATGG 1446  
| | | | | | | | | |  
Db 1438 CAAATGCTCCGGATATGG 1456  
| | | | | | | | | |

RESULT 12  
BC012075 2593 bp mRNA linear PRI 06-AUG-2001  
LOCUS  
DEFINITION Homo sapiens, hypothetical protein similar to beta-transducin  
family, clone MGC:19970 IMAGE:4652713, mRNA, complete cds.  
ACCESSION BC012075  
VERSION BC012075.1 GI:15082334  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 2593)  
Strausberg,R.  
Direct Submission  
Submitted (02-AUG-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 29 Row: 0 Column: 21  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7022501.  
Location/Qualifiers  
1. .2593  
/organism="Homo sapiens"  
/db\_xref="LocusID:54475"  
/db\_xref="taxon:9606"  
/clone="MGC:19970 IMAGE:4652713"  
/tissue\_type="Placenta, choriocarcinoma"  
/clone\_lib="NIH\_MGC\_21"  
/lab\_host="DH10B-R"  
/note="vector: pOTB7"  
14. .1471  
/codon\_start=1  
/product="hypothetical protein similar to beta-transducin  
family"  
/protein\_id="AAH12075.1"  
/db\_xref="GI:15082335"  
/translation="MAAAVADEAVARDVQRLLVQFQDEGGQLLGGSPFDVPVDITPDRL  
QLVCNALLAQEDPLPLAFFVHDAEIVSSLGKTLQESQAVETEKVLDLIYQPAIFRVA  
VTRCTSSLEGHSEAVISVAFSPTKYLASGSDTVRFWDLSSTPHTCTCKGRHWVL  
SISWSPDGKKLASGCKNGOILLWDPSTGKQVGRTRLAGHSKWITGLSWEPLHANECPY  
VASSKDSGVRIWDTTAGCERILTGHQSVTCLRWGDDGLLYSASQDRTIKVWRAHD  
GVLCRTLQGHGWNTMALSTDYALRTGAFEPAEASVNPQDLQSGIQELKERALSRYN  
LVRGQPERLVSGSDDFTLFLWSPAEDKKPLRTMTGHOALINQVLFSPDSRIVASAF  
DKSIKLWDGRTGKYLASLRGHVAAVYQIAWSADSRLLVSGSSDSTLKVDVKAQKIAM  
DLPGHADEVYAVDWSPDGQRVASGGKDKLRIWR"

BASE COUNT	582 a	734 c	749 g	528 t
ORIGIN				
Query Match	30.7%; Score 477.4; DB 9; Length 2593;			
Best Local Similarity	59.8%; Pred. No. 1.3e-108;			
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;				
QY	54	GATACAGGGCGGCTCGTTTACACGGGCGAGGAAGCCGGCCGCCAATCGACCTGCCGGC	113	
Db	64	GCTAGTGCAGTTCAGGATGAGGGCGGCGAGCTGCTGGGTCCCGCTCGACCTGCCCGT	123	
QY	114	AGGAATCACTACCCAGCAATTGGGACTGATTGCAACGGCTGCTGAAAAACGAGGAAGC	173	
Db	124	GGACATCACCCCGGACAGGCTGCTGCAACGGCTACTGGCCACGAGGATCC	183	
QY	174	CACTCCATATTGTTTTTCGTGGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACAGTT	233	
Db	184	CCTGCCACTGGCTTTCTTTGTCCACGATGCTGAGATCGTCTCCTACCTGGGGAAGCGTT	243	
QY	234	GGACTTGGCGTCAGTGGACACCGAAACGTGATCGATATGTGTATCAGCCACAGCGGT	293	
Db	244	GGAGTCCCAGGAGTGGAGACAGAGAGGTCTCTAGACATCATCTACCAGCCACAGGCTAT	303	
QY	294	TTTCAAAGTGCGCCAGTGCACAAGATGCACGAGTTCCATGCCGGGACAGCCCGAGGCTGT	353	
Db	304	CTTCAGAGTCCGGGCTGTGACTCGCTGCACCAAGTCCCTTGGAGGGTCACAGTGAGGCAGT	363	
QY	354	GGTTTCGCTGAATTTACGCCCGGATGGTCTCATCTCGCCAGTGGGAAGTGGCGACACCAC	413	
Db	364	CATTCTGTGGCTTCAGCCCTACGGGAAGTACCTGGCCAGTGGCTCTGGAGACACCAC	423	
QY	414	AGTGGGATTGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCA	473	
Db	424	CGTGGCTTCTGGGATCTCAGCACAGAGACACCACATTTTCACATGCAAGGACACAGACA	483	
QY	474	GTGGTTCTGTGGGTATCCTGGGCTCCGGATGGCAACGGTGGCCAGCGTTGCAAAAGC	533	
Db	484	CTGGGTCTTTAGTATATCTCTGGTCTCCAGATGGCAAGAGTGGCTCAGGCTGCAAGAA	543	
QY	534	GGGCTCTATAATCATCTGGACCCGGAGACGGGTTCAGCAAGAGGGGGAGCCCTTGATGG	593	
Db	544	TGGCAGATTCTCTCTGGGACCCCAAGCACAGGGAAGAGGTGGGCAGGACCCCTCGCTGG	603	
QY	594	GCACAAGAAACACATCAACTGCCTCGCTGGGAACCGTATCATCGCGATCCGGAGTGCAG	653	
Db	604	CCACAGCAAGTGGATCACAGGCTGAGCTGGGAGCCCTCCATCGGAACCTGAGTGGCG	663	
QY	654	GAACTTGTTCGGCCAGTGGAGACGGGGACTGCCGGATTGGGACGTAAATTTGGGCCA	713	
Db	664	CTATGTGGCCAGCAGCTCCAAGGATGGCAGTGTGCGGATCTGGGACACACAACTGCAGG	723	
QY	714	GTGCTTATGAACATTCGCGGACACACAAATGCTGTGACAGCAGTGAAGTGGGTGGAGC	773	
Db	724	CTGTGAGCGCATCTCACCGGGACACACCCAGTCCGCTCACCTGTCTCCGGTGGGAGGGA	783	
QY	774	GGGCTTATTTATACATCTCCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG	833	
Db	784	CGGGCTTCTCTACTCTGCCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG	843	
QY	834	AATCTTGTGCGGACGTTCTCTGGGCCAAGCTCACTAGGTTAAACAACATTCGGCTGAGCAC	893	
Db	844	TGTGCTGTGCGGACTCTGCAAGGCCACCGGCCACTGGGTGAACACCATGGCCCTCAGCAC	903	
QY	894	CGATTACGTCCTGGCACCTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCC---	950	
Db	904	TGACTATGCCCTGGCACCTGGGGCCCTTTGAACCTGCTGAGGCCTCAGTTAATCCCCAAGA	963	
QY	951	CCTCAGTTTGAGCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTG	1010	
Db	964	CCTCAAGGATCTCTGCAGGAGTTGAAGGAGAGGGGCTCTGAGCCGATACAACCTCGTGG	1023	
QY	1011	CCCTGACGAGGTGGAGTGGCTGGTTCTCTGTTGGATGACAACACCCCTCTATCTGTGG--	1068	



Db 1024 GGGCCAGGCTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTACCTTATTCCTGTGGTC 1083  
QY 1069 -CGGAACAACCAAGAGTCCGTTGAGCGCATGACAGGGCAGACAGAACGCTGGTCAACGA 1127  
Db 1084 CCCAGCAGAGGACAAAAGCCCTCTACTCGGATGACAGGACACCAAGCTCTCATCAACCA 1143  
QY 1128 TGTGAATATATTCGCCGGATGTAAAGCTAATTCGCTCTGCTTTCATTTGACAAAGTCAAGTGG 1187  
Db 1144 GGTGCTCTCTCTCCGACTCCCGCATCGTGGCTAGTGCCTCTTGTGACAAAGTCCATCAA 1203  
QY 1188 TCTGTGGGAGCCAGCGATGGTCACTACATGGCCACCTTCCGGGGTCAATGTGCAGGCTGT 1247  
Db 1204 GCTGTGGGATGGCAGGACGGGCAAGTACCTGGCTTCCCTACGGGGCCACGTGGCTGCCGT 1263  
QY 1248 TTACACGGTTCCTGTGCTCCGGACTCCCGCTTGATTTTTCGGGCAGCAAGAGACTCAAC 1307  
Db 1264 GTACCAGATGCGTGCAGTGCAGTCCGCTCTGGTCAAGGGCAGCACTGACAGCAC 1323  
QY 1308 TCTAAAGTATGGAGTGTGCAGACAGAAAGAACTGGCACAGGAGTGCCTGGACATGCCGA 1367  
Db 1324 ACTGAAGTGTGGGATGTGAAGGCCAGAAAGCTGGCCATGGACCTGCCCGCCACGCCGA 1383  
QY 1368 TGAGGTGTTCGGAGTGGACTGGGCCCCCGATGGCTCTAGAGTGGCTCTGGTGGCAAGGA 1427  
Db 1384 TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGATGGCAAGTGGTGGGAAGGA 1443  
QY 1428 CAAAGTTATAAGCTATGG 1446  
Db 1444 CAAATGCTCCGGATATGG 1462

RESULT 13  
AX405715 2638 bp DNA linear PAT 14-JUN-2002  
LOCUS Sequence 130 from Patent WO0222660.  
DEFINITION AX405715  
ACCESSION AX405715  
VERSION AX405715.1 GI:21438869  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A., Ren,F.,  
Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.  
TITLE Novel nucleic acids and polypeptides  
JOURNAL Patent: WO 0222660-A 130 21-MAR-2002;  
HYSEQ, INC. (US)  
FEATURES  
source Location/Qualifiers  
1..2638  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
CDS 82..1539  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD34822.1"  
/db\_xref="GI:21438870"  
/translation="MAAAVADEAVARDVQRLLVQFDEGGQLGSPFDVPVDTDPRL  
QLVCNALLAQEDPLPLAFFVHDAEIVSSLGKTLSEQAVETEKVLDIYQPOAIFRVA  
VTRCTSSLEGHSEAVISVAFSPTGKYLASGGDTFVRFWDLSTPHTFCKGHRHWL  
SISWSPDGKKLASCKNGQILLWDPSTGKQVGRITLAGHSHKWTGLSWEPLHANPECRY  
VASSKDGSVRIWDTTAGRCERILTGHTQSVTCLRWGGDGLLYSASQDRITKVMRAHD  
GVLCRTLQGHGHVWNTMALSTDYALRTGAFEPAEASVNPQDLQSLQELKERALSRYN  
LVRGQGPRLVSGDDFTFLWSPAEDKKPLTRMTGHQALINQVLFSPDSRIVASAF  
DKSIKLWDGRTGKYLASLRGHVAAVQIAWSADSKLLVSGSSDSTLKVWDVKAQKIAM  
DLPGHADEVYAVDWSPPQORVASGGKDKCLRWRR"  
BASE COUNT 566 a 767 c 764 g 541 t  
ORIGIN  
Query Match 30.7%; Score 477.4; DB 6; Length 2638;  
Best Local Similarity 59.8%; Pred. No. 1.3e-108;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 54 GATACAGGCGGCGCTCGTTTACACGGGGGAGGAAGCGCGCCCGCAATCGACCTGGCGGC 113  
Db 132 GCTAGTGCAGTCCAGGATGAGGGCGGCAGCTGCTGGGTTCCCGCTCGACGTCGCCGT 191  
QY 114 AGGAATCACTACCAGCAATTTGGACTGATTTTGAACGCGCTGCTGTAANAACGAGAAGC 173  
Db 192 GGACATACCCCGGACAGGCTGCAGCTCGTGTGCAACGCGCTACTGGCCCCAGGAGATCC 251  
QY 174 CACTCCATATTTGTTTTTCTGGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACAGCTT 233  
Db 252 CCTGCCACTGGCTTTCTTTGTCCAGGATGCTGAGATGCTCTCTCACTGGGGAACAGCTT 311  
QY 234 GGACTTGGCGTCACTGGACACCGAAACCTGATCGATATTTGTGATCAGCCACAGGCGGT 293  
Db 312 GGAGTCCCAGGAGTGGAGACAGAGAGGTCTTAGACATCATCTACAGCCACAGGCTAT 371  
QY 294 TTCAAAGTGGCGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCAGGCTGT 353  
Db 372 CTTCAAGTCCGGGCTGTGACTCGCTGCACAGCTCCTTGGAGGFTCACAGTGAGGCAGT 431  
QY 354 GGTTCGCTGAATTTACAGCCCGGATGGTGTCTCATCTGCCAGTGGAAAGTGGCAGACACC 413  
Db 432 CATTTCTGTGGCCTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACCAC 491  
QY 414 AGTCGATTTGTGGATCTTAACACAGACAGACACCGCACTTCACTGCACAGGTCATAAGCA 473  
Db 492 CGTGGCTTCTGGGATCTCAGACACAGACACACCATTTACATGCAAGGGACACAGACA 551  
QY 474 GTGGGTTCTGTGCGTATCCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCCAAGC 533  
Db 552 CTGGGTCTTAGTATATCTCTGGTCTCCAGATGGCAAGAGCTGGCCTCAGGCTGCAAGAA 611  
QY 534 GGGCTCTATAATCATCTGGACCCCGAGACGGGTACAGAGAGGGCGGACCTTGAGTGG 593  
Db 612 TGGCCAGATTTCTCTGGGACCCCAAGCACAGGAGCAGGTGGCAGGACCCCTCGCTGG 671  
QY 594 GCACAAGAAACACATCAACTGCTCGCTGGGAAACCGTATCATCCGATCCGGAGTCCAG 653  
Db 672 CCAAGCAACTGGATCAAGGCTGAGTGGAGCCCTCCATGGCAACCTGAGTGGCG 731  
QY 654 GAAACTTGCTTCCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAAATTTGGGCCA 713  
Db 732 CTATGTGGCCAGCAGCTCCAAGGATGSCAGTGTGCGGATCTGGGACACAACTGCGGCG 791  
QY 714 GTGCCTTATGAACATTTGCCGGACACACAAATGCTGTGACAGCAGTGGGTGGAGC 773  
Db 792 CTGTGAGCGCATCTCACGGGGCACACCCAGTCCGTGCTACCTGTCTCCGGTGGGAGGGA 851  
QY 774 GGGCCTTATTATACATCTCCAAAGATGCGACAGTGAAGATGTGGCGAGCAGCTGATGG 833  
Db 852 CGGGCTTCTCTACTCTGCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG 911  
QY 834 AATCTTGTGCGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAAATTTGGCTGAGCAC 893  
Db 912 TGTGCTGTGCGGACTCTGCAAGGCCACGGCCACTGGGTGAACACCATGGCCCTCAGCAC 971  
QY 894 CGATTACGCTCTGCGCACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCC--A 950  
Db 972 TGACTATGCCCTGCGCACTGGGGCCCTTGAACCTGCTGAGGCCCTCAGTTAATCCCCAAGA 1031  
QY 951 CCTCAGTTTGAGCACTGAGGAATTGCAGGAATCTGCTTGAAGCGCTACCAGGCGGTGTG 1010  
Db 1032 CCTCCAAGGATCTTGCAGGAGTTGAAGGAGAGGGTCTGAGCCCATACAACTCGTGGC 1091  
QY 1011 CCCTGACGAGGTGGAGTGGTGGTTTCTGTTCCGATGACAAACACCTCTATCTGTGG-- 1068  
Db 1092 GGGCCAGGTTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACTTATTCCTGTGGTC 1151  
QY 1069 -CGGAACAACCAAGAACAGTGGTTGAGCGCATGACAGGGCAGACCAACGTTGGTCAACGA 1127  
Db 1152 CCCAGCAGAGGACAAAAAGCCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA 1211  
QY 1128 TGTGAATATTCGCCGGATGTAAAGCTAAATTCGCTGTCTTCAATTTGACAAGTCAAGTGG 1187



Db 1212 GGTGCTCTTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCTTGGACAAAGTCCATCAA 1271  
QY 1188 TCTGTGGGAGCCAGCGATGGTCAGTACATGCCACCTTCCGGGGTCATGTGCAGGCTGT 1247  
Db 1272 GCTGTGGGATGGCAGGACGGGCAAGTACCTGGCTTCCCTACCGGGCCACGTGGCTGCCGT 1331  
QY 1248 TTACACGGTTGCCGTGGTCCGGGACTCCCGCTTGATTGTTTCCGGCAGCAAGACTCAAC 1307  
Db 1332 GTACACAGATTGGGTGGTCAGCTACAGTCGGCTCCTGGTCAGCGGCAGCAGTGACAGCAC 1391  
QY 1308 TCTAAAGTATGGAGTGTGCAGACGAAGAACTGGCACAGGAGCTGCCCTGGACATGCGGA 1367  
Db 1392 ACTGAAGGTGTGGGATGTGAAGGCCAGAAAGCTGGCCATGGACCTGCCCGGCCACGCGGA 1451  
QY 1368 TGAGGTGTTCCGGAGTGGACTGGGGCCCGATGGCTCTAGAGTTGCCCTCTGRTGGCAAGGA 1427  
Db 1452 TGAGGTATATGCTGTTGACTGGAGTCCAGATGGCCAGAGTGGCAAGTGGTGGGAAGGA 1511  
QY 1428 CAAAGTTATAAAGCTATGG 1446  
Db 1512 CAAATGCCCTCCGGATATGG 1530

RESULT 14  
BC018399  
LOCUS  
DEFINITION Mus musculus, hypothetical protein similar to beta-transducin family, clone MGC:25690 IMAGE:3491925, mRNA, complete cds.  
ACCESSION BC018399  
VERSION BC018399.1 GI:17390942  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1784)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunnaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 30 Row: b Column: 14.  
Location/Qualifiers  
1. .1784  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="CZECH II"  
/clone="MGC:25690 IMAGE:3491925"  
/tissue\_type="Mammary tumor metastatized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMV."  
/clone\_lib="NCI\_CGAP\_Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

FEATURES  
source

CDS  
31. .1488  
/codon\_start=1  
/product="hypothetical protein similar to beta-transducin family"  
/protein\_id="AAH18399.1"  
/db\_xref="GI:17390943"  
/db\_xref="LocusID:54475"  
/translation="MAAAVVEEAAAAGDVORLLVQFQDEGQLLGSPFDVVPVDITPDKL QLVCNALLAQEEPLPLAFYVHDAEIVSSLGKLTLESQSVETEKIVDIIYQPQAVFRVRA VTRCTSSLEHSEAVISVAFSPTGKYLASGSDTTVRFWDLSTETPHFTCKGHRHWVL SISWSPDGKKLASGCKNGQILLNDPSTGLQVGRTLTGHSKWITLSWEPHLMNPECRY VASSSKDGSVRVWDTTAGCERILITGHTQSVTCLRWGGDGLLYSASQDRTIKVWRAHD GVLCRTLQHGHWVNTMALSTDYALRTGAPEAEATVNAQDLQGLSKELKERASSRYN LVRGQGPRLVSGSDDFTLFWSPAEDKKPLARMTGHQALINQVLFSPDSRIVASASE DKSIKLWDGRTGKYLASLRGHVAAVQIAWSADSRLLVSGSSDSTLKVWDVKAQKLAT DLPGHADEVYAVDWSPDGQRVASGGKDKCLRIWRR"  
BASE COUNT 397 a 521 c 539 g 327 t  
ORIGIN  
Query Match 30.6%; Score 476.6; DB 10; Length 1784;  
Best Local Similarity 59.2%; Pred. No. 1.9e-108;  
Matches 850; Conservative 0; Mismatches 579; Indels 6; Gaps 2;  
QY 18 GGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCGCTCGTTTACAC 77  
Db 45 GGTGAAGAGCGCGCGCAGCGACGTGTCAGCGGCTGCTCGTACAGTCCAGGATGAGGG 104  
QY 78 GGGCGAGGAAGCCGGCCGCCAATCGACCTGCCGCGAGGAATCACTACCCAGCAATTGGG 137  
Db 105 CGGTCACTGCTGGGCTCCCCGTTTCGACGTGCCCGTGGACATCACGCCGACAAAGCTGCA 164  
QY 138 ACTGATTTGCAACGGCGCTGCTGAAAACAGCAGGAAGCCACTCCATATTTGTTTCGTGGG 197  
Db 165 GCTGGTGTGCAACGGCGCTGCTGGCCCGCAGGAGGAGCCCTGCCCTTCTATGTCCA 224  
QY 198 CGAGGATGAGATCAAGAAGAGCCCTGGAGGACACAGTTGGAGTGGCGTACGTGGACACCGA 257  
Db 225 CGATGCTGAGATTGTGTCCTCCCTGGGGAGAGACGCTGGAGTCCCAGTCTGTGGAGACAGA 284  
QY 258 AAACGTGATCGATATTGTGTATCAGCCACAGCGCGGTTTTCAAAGTCCGCCCAGTGACAAG 317  
Db 285 GAAATCGTCGATATCATCTACCAGCCCCCAAGCTGTCTTCAGGGTCCGGGCTGTGACGG 344  
QY 318 ATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTGAATTTTCAGCCCGGA 377  
Db 345 GTGCACACGCTCATTTGAGGGTTCACAGCGAAGCTGTCAATTTCTGTGGCTTCAGCCCCAC 404  
QY 378 TGGTGTCTCATCTCGCCAGTGAAGTGGCGACACCAACAGTGCAGTGTGGGATTTGGGATCTTAACAC 437  
Db 405 AGGAAAGTACCTGGCCAGTGTGTTCCGGAGACACCACTGTGCGCTTCTGGGATCTCAGCAC 464  
QY 438 AGAGACACCGCACTTCACCTGCACAGGTTCATAAGCAGTGGGTTCTGTGCGTATCTCTGGGC 497  
Db 465 TGAGACACCGCACTTCACATGCAAAAGGACACCGGCACCTGGGTCTCTAGCATATCTGTGTC 524  
QY 498 TCCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAAAGCGGGCTCTATAATCATCTCTGGGACC 557  
Db 525 CCCAGATGGCAAGAAACTGGGCTCAGGCTCAGGCTGCAAGAAATGGCCAGATTTCTGCTGTGGGACC 584  
QY 558 GGAGACGGGTGACAGAGAAGGGCGACCCCTTGAGTGGGCAAGAAACACATCAACTGCCT 617  
Db 585 AAGCACGGGGTGCAGGTGGGAGGACCCCTCACTGGCCACAGCAAAATGGATCAGAGGCT 644  
QY 618 CGCCTGGGAACCGTATCATCGCGATCCGGAGTCCGAGTGCAGAAACTTGTCTCCGCCAGTGGAGA 677  
Db 645 GAGCTGGAGAGCCCTTCACATGAACCCCGAGTGCCTACGTGGCCAGCAGCTCCAAAGA 704  
QY 678 CGGGGACTGCCGGATTTGGGAGCTAAAATTGGGCCAGTGCCTTATGAACATTGCCGGACA 737  
Db 705 CGGCAGTGTACGAGTCTGGGATACAAACCGCAGGCCGCTGTGTAACGCATCTCTCACGGGACA 764  
QY 738 CACAAATGCTGTGACAGCAGTACAGATGGGTGGAGCGGGCTTATTTATACATCTCTCCAA 797

Db 765 CACGCAGTCAGTCACCTGCCTCCGATGGGGAGGAGATGGGCTTCTATATATCTGCTTCCCA 824

QY 798 AGATGCGCAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGCGGAGACGTTCTCTGG 857

Db 825 GGACCGAACCATCAAAGTCTGGAGGGCCCATGACGGTGTGCTGTGCTACTCTTCAAGG 884

QY 858 CCAAGCTCACTGGGTAAACAACATTTGGCTGAGCACCAGATTACGTCCTGCGCACTGGTCC 917

Db 885 TCATGGCCACTGGTGAACACCATGGGCTTAGCACAGACTATGCCCTGCGCACAGGGG 944

QY 918 ATTCCATCCGGTGAAGGATCGCTCCAAGAGCC--ACCTCAGTTTGAGCACTGAGGAATT 974

Db 945 CTTTGAGCCTGCTGAAGCCACAGTGAATGCCCAAGACCTGCAGGGGTCCTTGRAAGAGCT 1004

QY 975 GCAGGAATCTGCCTTGAAGCGCTACCAAGGCCGTGTGCCCTGACGAGGTGGAGTCGCTGGT 1034

Db 1005 GAAGGAGAGGGCATCGAGCCGCTACAACCTCGTGGGGGCCAAGSCCCAGAGAGGCTGGT 1064

QY 1035 TTCTGTTCGGATGACAACACCCCTCTATCTGTGG--CGGAACAACCAAGAAAGTCCGT 1091

Db 1065 CTCTGGCTCAGACGACTTCACCTTATCTGTGTGCTGCCAGCAGAGGACAAGAGCCCT 1124

QY 1092 TGAGCGCATGACAGGCAACCAAGCGTGTCAACGATGTGAATATTCGCCGGATGATAA 1151

Db 1125 CGCCCGGATGACAGGACACCAAGCCCTCATCAACAGGTGCTCTCTCTCTGACTCGCG 1184

QY 1152 GCTAATTCGGTCTGCTTCATTGACAAGTCAGTGGGCTGTGGCGAGCCAGCGATGGTCA 1211

Db 1185 CATCGTGGCCAGTGCTCTTCGACAAGTCTATCAAGCTGTGGGATGGCCGACGGCAA 1244

QY 1212 GTACATGGCCACCTTCCGGGTCTATGTGCAGGCTGTATTACAGGTTGCCCTGGTCCGCGGA 1271

Db 1245 GTACCTAGCTTCCCTGCGTGGCCACGTGGCAGCGGTGTACCAAGATTGCCGTGCTCAGCTGA 1304

QY 1272 CTCCCGCTTGATTTCCGGCAGCAAGACTCAACTCTAAAGTATGGAGTGCAGAC 1331

Db 1305 CAGCCGACTATTAGTCAGGGCAGCAGTGCAGCAGCTGAAGGTGTGGGATGTGAAGGC 1364

QY 1332 GAAGAACTGGCAGCAGGAGCTGCTTGACATGCGGATGAGGTTCGGAGTGGACTGGGC 1391

Db 1365 CCAGAACTGGCCACGGACCTGCTGGCCATGGCATGGGATGAGGTATGCTGTGACTGGAG 1424

QY 1392 GCCCGATGGCTCTAGAGTTGCCCTCTGTTGGCAGGACAAAGTTATAAGCTATGG 1446

Db 1425 TCCAGATGCCAGAGAGTAGCAAGCGCGGGGAAGACAAGTGCCTCCGGATATGG 1479

RESULT 15  
AX135807  
LOCUS  
DEFINITION  
Sequence 1 from Patent WO0132614.  
ACCESSION  
AX135807  
VERSION  
AX135807.1 GI:14272042  
KEYWORDS  
human.  
SOURCE  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1928)  
AUTHORS  
Utku,N.  
TITLE  
Novel genes tzap7/a, tzap7/b and tzap7 involved in t cell  
activation and uses thereof  
JOURNAL  
Patent: WO 0132614-A 1 10-MAY-2001;  
Utku, Nalan (DE)  
FEATURES  
Source  
Location/Qualifiers  
1..1928  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
92..1546  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39574.1"  
/db\_xref="GI:14272043"

/translation="MVWIADAEAVARDVQRLIVQFQDEGGQLIGSPFVVPVDTTPDRLO  
LVCNALLAQEDPLSLAFFVHDAEIVSSLGKTESQAVETEKVLDIILPTQAVFKVRV  
TRCTSSLEGHTEAVISVAFSPTGKYLASGSDTTRFVWDLSTETPHFTSKGHTHWVLS  
IAWSPDGKKLASGCKNSQIFIWDPSTGKIQGKPLTGHSKWITWLCWEPLHNPESRYL  
ASASSGRVDRWDTAGTCERILTGHQSVTCLRWGGDGLIYSASQDRITIKVWRAHDG  
VLCRTLOGHGHWNMTALSTDYALRTGAFEPAEASVNPQDLQSLQELKERALSRYNL  
VRGQPERLVSGSDDFTLFLWSPAEDKKPLRTMTGHQALINQVLFSPDSRIVASASFD  
KSIKLWDGRTKYLASLRHVAAYQIAWSADSRLLVSGSSDSTLKVWDVKAQKLA  
LPGHADEVYAVDWSPDGQRVASGGKDKCLRIWRR"

BASE COUNT 432 a 540 c 550 g 406 t

ORIGIN

Query Match 30.0%; Score 466.2; DB 6; Length 1928;  
Best Local Similarity 59.3%; Pred. No. 7.7e-106;  
Matches 830; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

QY 54 GATACAGGCGCGCTCGTTTACACGGGCGAGGAAGCGGCGCCGCCAATCGACCTGCGCGC 113

Db 139 GCTAGTCAGTTCAGGATGAGGGCGGCGAGCTGCTGGGTTCCCGCTCGACGTGCCCGT 198

QY 114 AGGAATCACTACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAACACGAGGAAGC 173

Db 199 GGACATCACCCCGGACAGGCTGCAGCTGCTTTGCAACGCGCTACTGGCCCCAGGAGATCC 258

QY 174 CACTCCATATTTGTTTTCGTGGGCGAGGATGAGATCAAGAAGAGCTGGAGGACACGTT 233

Db 259 CCTGTCACTGGCTTCTTGTGCCAGATGCTGAGATCGTTCTCTCACTCGGAAAGACGTT 318

QY 234 GGACTTGGCGTCAGTGCACACCCGAAACAGTGCATCGATATTGTGATCATAGCCACAGCGGCT 293

Db 319 GGAGTCCAGGCACTGGAGACAGAGAGGTCTCTAGACATCATTTTACCCACACAAGCTGT 378

QY 294 TTTCAAAGTGGCCCCAGTGACAAGATGCACAGAGTTCATGCCGGGACACGCCGAGGCTGT 353

Db 379 GTTCAAAGTTCGTGCTTAACACGATGTACCAGCTCATTTGGAGGACACACCCGAGGCTGT 438

QY 354 GGTTCGCTGAATTCAGCCCCGGATGGTGTCTCATCTGCCAGTGGAAAGTGGCGACACCAC 413

Db 439 TATTTCACTAGCTTTCAGCCCCAAGTGAAGTATTTGGCAAGTGTCTTGGGACACTAC 498

QY 414 AGTGCATGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCA 473

Db 499 AGTCCGCTTTTGGGATCTCAGCACAGAAAACCTCCACATTTTACATCTAAAGGGCATACAC 558

QY 474 GTGGTTCGTGCGCTATCCTGGGCTCCGGATGGCAAAACGGTTGGCAGCGGTTGCCAAAGC 533

Db 559 CTGGGTTCTCAGTATGTGTTGGTCTCAGATGGCAAAAACCTTGCTCTCAGGATGTAAAAA 618

QY 534 GGGCTCTATAATCATCTGGGACCGGAGACGGGTTCAGCAGAGGGGCGACCCCTTGAGTGG 593

Db 619 TAGTCAGATCTTCAATTTGGGACCCCAAGCACAGGGAAGCAGATTGGCAAAACCATTAACAG 678

QY 594 GCACAAGAAACACATCAACTGCCTCGCCTGGGACCCGTATCATCTCGGATCCGGAGTGCAG 653

Db 679 GCACTCAAAGTGGATTACATGGCTGTGTTGGGAACCTCTCCACCTGAACCCAGAGAGCG 738

QY 654 GAACTTGTCTCCGCGAGTGGAGACGGGACTGCCGGATTTGGGACGTAATAATTTGGGCCA 713

Db 739 ATACCTAGCCAGTGCCTCCAGCGGCGCGCTCGACCGGATCTGGGACACAACTGCAGGCG 798

QY 714 GTGCCTTATGAACATTCGCGGACACACAAATGCTGTGACAGAGTGAAGTGGGTGGAGC 773

Db 799 CTGTGAGCGCATCTCACCGGGCACACCCAGTCCGCTCACCTGTCTCCGGTGGGAGGGGA 858

QY 774 GGGCCTTATTTATACATCTCCAAAGATCGCAGATCGAGATGTGGCGAGCAGCTGATGG 833

Db 859 CGGGCTTCTCTACTCTGCCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG 918

QY 834 AATCTTGTCCCGACGTTCTCTGCGCCAAAGCTCACTGGGTAAACAACATTTGGCTGAGCAC 893

Db 919 TGTGCTGTCCCGGACTCTGCAAGGCCACCGGCCACTGGGTGAACACCATGGCCCTCAGCAC 978

QY 894 CGATTACGTCCTGCGCACTGGTTCCTATTCATCCCGGTGAAGGATCGCTCCAAGAGCC--A 950

[illegible]





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 03:05:19 ; Search time 378 Seconds  
(without alignments)  
9264.173 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattcccaaaaaatgcagga.....aaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1555	100.0	1555	21 AAA27739	Drosophila Notchle
2	1487.8	95.7	1992	23 ABL05339	Drosophila melanog
3	1285.8	82.7	4164	23 ABL05338	Drosophila melanog
4	601.2	38.7	6223	23 ABL05232	Drosophila melanog
5	588.2	37.8	4141	23 ABL05214	Drosophila melanog
6	478	30.7	1859	22 AAF85075	Nucleotide sequenc
7	477.4	30.7	1853	22 AAH14166	Human cDNA sequenc
8	477.4	30.7	1854	22 AAF58316	Human GTP-binding
9	477.4	30.7	2638	24 ABN59719	Novel human coding

10	466.2	30.0	1928	22 AAF85073	Nucleotide sequenc
11	333.4	21.4	1726	21 AAC48725	Arabidopsis thalia
12	331.8	21.3	1643	21 AAC39906	Arabidopsis thalia
13	249.4	16.0	1170	22 AAF85074	Nucleotide sequenc
14	176	11.3	399	21 AAH30571	Human colon cancer
15	164	10.5	547	22 AAH05727	Human cDNA clone (
16	155.8	10.0	1722	20 AAX30340	DNA encoding a hum
17	153.6	9.9	573	22 AAX98017	Murine 7-transmemb
18	123.8	8.0	540	21 AAF07777	Fusarium venenatun
19	120.4	7.7	378	22 AAF67530	Novel human polynu
20	80.4	5.2	316	21 AAF11429	Aspergillus niger
21	79.4	5.1	2984	22 AAK94739	Human full-length
22	79.4	5.1	3025	22 ABA06363	Human cDNA SEQ ID
23	78.4	5.0	300	20 AAZ14360	Human gene express
24	78.4	5.0	764	22 AAK91829	Human cDNA 5'-end
25	78.4	5.0	764	22 AAK93229	Human cDNA clone r
26	78.4	5.0	1950	22 AAK52860	Human polynucleoti
27	69	4.4	1925	22 AAK51876	Human polynucleoti
28	68.4	4.4	279	23 ABV55135	Human prostate exp
29	68.4	4.4	2977	24 ABQ72680	Human MDDT encodin
30	68.2	4.4	1464	24 ABN59758	Novel human coding
31	68.2	4.4	1751	22 AAF58329	Human GTP-binding
32	68.2	4.4	2064	21 AAF22375	Human secreted pro
33	64	4.1	1591	23 ABL23519	Drosophila melanog
34	62.6	4.0	462	23 ABV57002	Human prostate exp
35	62.2	4.0	559	23 ABV54541	Human prostate exp
36	62	4.0	1810	21 AAC77522	Human ORFX ORF3077
37	61.8	4.0	2359	15 AAK70726	TATA-binding prote
38	61.4	3.9	821	22 AAK92334	Human cDNA 5'-end
39	61.4	3.9	821	22 AAK93787	Human cDNA clone r
40	61.4	3.9	1890	22 AAK94610	Human full-length
41	60.2	3.9	2359	17 AAT42212	Drosophila TATA-bi
42	60.2	3.9	2359	18 AAT79603	TATA-binding prote
43	60.2	3.9	2826	23 ABL13261	Drosophila melanog
44	60.2	3.9	3965	23 ABL11658	Drosophila melanog
45	60.2	3.9	4826	23 ABL13260	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAA27739  
ID AAA27739 standard; cDNA; 1555 BP.

XX AAA27739;

XX 29-AUG-2000 (first entry)

DE Drosophila Notchless (Nle) cDNA.

KW Notchless; Nle gene; Notch; signalling; neurodegenerative disease;  
KW cancer; diagnosis; cytostatic; neuroprotective; gene therapy; ss.

OS Drosophila melanogaster.

XX Key Location/Qualifiers  
FT CDS 13..1455  
FT /tag= a  
FT /transl\_except= (pos:859..861, aa:His)

XX WO200026364-A1.  
PN 11-MAY-2000.  
XX 03-NOV-1999; 99WO-IB01891.  
XX 03-NOV-1998; 98GB-0024045.  
XX (EUMO-) EURO MOLECULAR BIOLOGY LAB.  
XX Cohen S, Bouwmeester A, Royet J;  
XX

DR WPI; 2000-365613/31.  
DR P-PSDB; AAY79678.  
XX Novel Notchless protein and nucleic acids encoding them useful for  
PT treating and preventing cancer and neurodegenerative diseases -  
PT  
XX  
PS Claim 19; Page 44-45; 52pp; English.  
XX  
CC The present sequence is that of cDNA encoding Notchless, a novel  
CC protein of Drosophila. The cDNA was isolated from an eye disc cDNA  
CC library by screening with a genomic DNA fragment that flanks the P  
CC element insertion site in Drosophila 1(2)AAK13714. Notchless was  
CC identified in a screen for dominant modifiers of a Notch mutant  
CC phenotype in the Drosophila wing. The mutant dominantly suppressed  
CC the wing notching phenotype of notchoid mutations, and the Notchless  
CC protein was shown to bind to the cytoplasmic domain of Notch.  
CC Notchless modifed Notch signalling activity in a variety of  
CC Notch-dependent signalling process in both Drosophila and Xenopus  
CC embryos. The Notchless protein, and nucleic acids encoding it, can  
CC be used in methods for the diagnosis and therapy of certain diseases,  
CC particularly cancer and neurodegenerative diseases (claimed). A  
CC probe capable of screening for the Notchless gene (Nle), a cloning  
CC or expression vector comprising Notchless DNA, cDNA or RNA, a host  
CC cell, a transgenic animal, and a method for production of Notchless  
CC protein in the host cell are also claimed.  
XX  
SQ Sequence 1555 BP; 418 A; 373 C; 448 G; 316 T; 0 other;  
Query Match 100.0%; Score 1555; DB 21; Length 1555;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AATTCCCAAAAATGCAGGAGACGGACACGGGAGCAAGAGGCCACGCCACATACGATACAG 60  
Db 1 AATTCCCAAAAATGCAGGAGACGGACACGGGAGCAAGAGGCCACGCCACATACGATACAG 60  
QY 61 GCGCGCCTCGTTTACACGGCGGAGGAAGCGCGCCGCCAATCGACCTGCCGGCAGGAATC 120  
Db 61 GCGCGCCTCGTTTACACGGCGGAGGAAGCGCGCCGCCAATCGACCTGCCGGCAGGAATC 120  
QY 121 ACTACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCA 180  
Db 121 ACTACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCA 180  
QY 181 TATTTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTG 240  
Db 181 TATTTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTG 240  
QY 241 GCGTCAGTGGACACCGAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAA 300  
Db 241 GCGTCAGTGGACACCGAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAA 300  
QY 301 GTCGCCCGCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCG 360  
Db 301 GTCGCCCGCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCG 360  
QY 361 CTGAATTTACCCCGGATGGTGCTCATCTCGCCAGTGGGAAGTGGCGACACACAGTGGGA 420  
Db 361 CTGAATTTACCCCGGATGGTGCTCATCTCGCCAGTGGGAAGTGGCGACACACAGTGGGA 420  
QY 421 TTGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCAGTGGGTT 480  
Db 421 TTGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCAGTGGGTT 480  
QY 481 CTGTGCGTATCCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAAAGCGGGCTCT 540  
Db 481 CTGTGCGTATCCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAAAGCGGGCTCT 540  
QY 541 ATAATCATCTGGGACCCGGAGACGGGTCAGCAGAAGGGSGGACCCCTTGAGTGGGCACAAG 600  
Db 541 ATAATCATCTGGGACCCGGAGACGGGTCAGCAGAAGGGSGGACCCCTTGAGTGGGCACAAG 600  
QY 601 AAACACATCAACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTT 660

Db 601 AAACACATCAACTGCCTCGCCTGGAAACCGTATCATCGCGATCCGGAGTGCAGGAAACTT 660  
QY 661 GCTTCGCCCAGTGGAGACGGGACTGCCGGATTGGGACGTAAAAATTGGGCCAGTGCCCTT 720  
Db 661 GCTTCGCCCAGTGGAGACGGGACTGCCGGATTGGGACGTAAAAATTGGGCCAGTGCCCTT 720  
QY 721 ATGAACATTGCCGGACACACAAATGCTGTGACAGCAGTGGGGTGGAGCGGGCCCTT 780  
Db 721 ATGAACATTGCCGGACACACAAATGCTGTGACAGCAGTGGGGTGGAGCGGGCCCTT 780  
QY 781 ATTTATACATCCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTG 840  
Db 781 ATTTATACATCCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTG 840  
QY 841 TGCCGGACGTTCTCTGGCCAAGCTCACCTGGGTAAACACATTCGCTGAGCACCGGATTAC 900  
Db 841 TGCCGGACGTTCTCTGGCCAAGCTCACCTGGGTAAACACATTCGCTGAGCACCGGATTAC 900  
QY 901 GTCCTGCCCACTGGTCCATTCCATCCGCTGAAGGATCGCTCCAAGAGCCACCTCAGTTTG 960  
Db 901 GTCCTGCCCACTGGTCCATTCCATCCGCTGAAGGATCGCTCCAAGAGCCACCTCAGTTTG 960  
QY 961 AGCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACAGGCCGCTGTCCTGACGAG 1020  
Db 961 AGCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACAGGCCGCTGTCCTGACGAG 1020  
QY 1021 GTGGAGTCGCTGGTTTCCCTGTCGGATGACAAACACCCCTCTATCTGTGGCGAACAACCCAG 1080  
Db 1021 GTGGAGTCGCTGGTTTCCCTGTCGGATGACAAACACCCCTCTATCTGTGGCGAACAACCCAG 1080  
QY 1081 AACAAAGTCGCTGAGCGCATGACAGGGCACAGAAACCGTGGTCAACGATGTGAAATATTTCG 1140  
Db 1081 AACAAAGTCGCTGAGCGCATGACAGGGCACAGAAACCGTGGTCAACGATGTGAAATATTTCG 1140  
QY 1141 CCGGATGTAAAGCTAATTGCGTCTGCTTCAATTGACAAAGTCAGTCCGTGTGGCGAGCC 1200  
Db 1141 CCGGATGTAAAGCTAATTGCGTCTGCTTCAATTGACAAAGTCAGTCCGTGTGGCGAGCC 1200  
QY 1201 AGCGATGGTCAGTACATGGCCACCTTCCGGGGTCATGTGACGGCTGTTTACACGGTTGCC 1260  
Db 1201 AGCGATGGTCAGTACATGGCCACCTTCCGGGGTCATGTGACGGCTGTTTACACGGTTGCC 1260  
QY 1261 TGGTCCGGGACTCCCGTTGATGTTCCGGCAGCAAAAGACTCAACTCTAAAGTATGG 1320  
Db 1261 TGGTCCGGGACTCCCGTTGATGTTCCGGCAGCAAAAGACTCAACTCTAAAGTATGG 1320  
QY 1321 AGTGTGCAGACGAAGAAACTGGCACAGGAGTGCCTGGACATCGGGATGAGGTGTTCCGGA 1380  
Db 1321 AGTGTGCAGACGAAGAAACTGGCACAGGAGTGCCTGGACATCGGGATGAGGTGTTCCGGA 1380  
QY 1381 GTGGACTGGGCCCGCGATGGCTCTAGAGTTGCCTCTGGTGCAAGGACAAAGTTATAAAG 1440  
Db 1381 GTGGACTGGGCCCGCGCGATGGCTCTAGAGTTGCCTCTGGTGCAAGGACAAAGTTATAAAG 1440  
QY 1441 CTATGGGCTTATTAAACAAATCATTAACCTGTACACGGTAAGAAAACTACTAGGAATAAAG 1500  
Db 1441 CTATGGGCTTATTAAACAAATCATTAACCTGTACACGGTAAGAAAACTACTAGGAATAAAG 1500  
QY 1501 TAAAACGTCCTGAGTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1555  
Db 1501 TAAAACGTCCTGAGTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1555

RESULT 2  
ABL05339  
ID ABL05339 standard; cdNA; 1992 BP.  
XX  
AC ABL05339;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10499.





QY 1504 AACGTCCTGAGTAA 1518  
Db 1519 AACGTCCTGAGTAA 1533

RESULT 3  
ABL05338

ID ABL05338 standard; cDNA; 4164 BP.  
XX  
AC ABL05338;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10496.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
PA  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR  
DR P-PSDB; ABB61235.  
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX

PS Claim 1; SEQ ID NO 10496; 21pp + Sequence Listing; English.  
XX

CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 4164 BP; 947 A; 1030 C; 1101 G; 1086 T; 0 other;

Query Match 82.7%; Score 1285.8; DB 23; Length 4164;  
Best Local Similarity 89.2%; Pred. No. 1.7e-279;  
Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;

QY 7 CAAAAAATGCAGGAGACGACACGGAGCAAGAGGCCACGCCACATACAGTACAGCGCGC 66  
Db 1019 CAGAAATGCAGGAGACGACACGGAGCAAGAGGCCACGCCACATACAGTACAGCGCGC 1078

QY 67 CTCGTTT---ACACGGGCGAGGAAGCCGGCCCGCCCAATCGACCTGCCGGCAGGATCACT 123  
Db 1079 CTCGTTTCGGACACGGGCGAGGAAGCGGGCCCGCCCAATCGACCTGCCGGCAGGATCACT 1138

QY 124 ACCCAGCAATGGGACTGATTTGCAACGGCGTCTGTGAAACAGGAGGCCACTCCATAT 183  
Db 1139 ACCCAGCAATGGGCTGATTTGCAACGGCGTCTGTGAAACAGGAGGCCACTCCATAT 1198

QY 184 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 243

Db 1199 TTTGTTTTTCGTGGCGAGGATGAGATCAAGAAAGAGCCTGGAGGACACGTTGGCGG 1258

QY 244 TCAGTGGACACCCGAAAACGTCGATTCGATATGTGTATCAGCCACAGGGCGTTTTCAAAGTG 303  
Db 1259 TCAGTGGACACCCGAAAACGTCGATTCGATATGTGTATCAGCCACAGGGCGTTTTCAAAGTG 1318

QY 304 CGCCCACTGACAAGATGCACGAGTTCATCCGCGGACACGCCGAGGCTGTGTTTCGCTG 363  
Db 1319 CGCCCACTGACAAGATGCACGAGTTCATCCGCGGACACGCCGAGGCTGTGTTTCGCTG 1378

QY 364 AATTTCAAGCCCGGATGGTGCTCATCTCGCCAGTGGAGTGGCGACACACAGTGGGATG 423  
Db 1379 AATTTCAAGCCCGGATGGTGCTCATCTCGCCAGTGGAGTGGCGACACACAGTGGGATG 1438

QY 424 TGGGATCTTAACACAGAGACACCGGACTTCACCTGCACAGGTCAATGAAGCAGTGGGTTCTG 483  
Db 1439 TGGGATCTTAACACAGAGACACCGGACTTCACCTGCACAGGTCAATGAAGCAGTGGGTTCTG 1498

QY 484 TGCCTATCCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAAAGCGGCTCTATA 543  
Db 1499 TGCCTATCCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAAAGCGGCTCTATA 1558

QY 544 ATCATCTGGGACCCCGSAGACGGGTCAGCAGAAAGGGGCGACCCCTTGAGTGGGCGACAAGAAA 603  
Db 1559 ATCATCTGGGACCCCGSAGACGGGTCAGCAGAAAGGGGCGACCCCTTGAGTGGGCGACAAGAAA 1618

QY 604 CACATCAAACTGCCTCGCCTGGGAACCGTATCATCCGGATCCGGAGTGCAGGAACCTTGCT 663  
Db 1619 CACATCAAACTGCCTCGCCTGGGAACCGTATCATCCGGATCCGGAGTGCAGGAACCTTGCT 1678

QY 664 TCCGCCAGTGGAGAGGGGACTGCGGGATTTGGGACGTAATAATTGGGCCAGTGCCTTATG 723  
Db 1679 TCCGCCAGTGGAGAGGGGACTGCGGGATTTGGGACGTAATAATTGGGCCAGTGCCTTATG 1738

QY 724 AACATGCCGGACACACAAAATGCTGTGACAGCAGTGAAGTGGGTTGGAGCGGCGCTTATT 783  
Db 1739 AACATGCCGGACACACAAAATGCTGTGACAGCAGTGAAGTGGGTTGGAGCGGCGCTTATT 1798

QY 784 TATACATCCTCCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTC 843  
Db 1799 TATACATCCTCCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTC 1858

QY 844 CGACGTTCTCTGGCCAAAGCTCACTGGGTAACAACAATTCGCTGAGCACCGATTACGTC 903  
Db 1859 CGACGTTCTCTGGCCAAAGCTCACTGGGTAACAACAATTCGCTGAGCACCGACTACGTC 1918

QY 904 CTGCGCACTGTGCCATTCCATCCGGTGAAGATCGCTCCAAGAGCCACCTCAGTTTGAGC 963  
Db 1919 CTGCGCACTGTGCCATTCCATCCGGTGAAGATCGCTCCAAGAGCCACCTCAGTTTGAGC 1978

QY 964 A-----C 965  
Db 1979 AGTAAGAAATTCATATGCTCTATATATGCTGTATATCCTAAAAGAGATTGTTTACAGC 2038

QY 966 TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCGCGGTGTCCTGACGAGGTGGA 1025  
Db 2039 TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCGCGGTGTCCTGACGAGGTGGA 2098

QY 1026 GTCGCTGGTTTCCTGTTCCGATGACAACACCCCTCTATCTGTGGCGGAACAACAGAACAA 1085  
Db 2099 GTCGCTGGTTTCCTGTTCCGATGACAACACCCCTCTATCTGTGGCGGAACAACAGAACAA 2158

QY 1086 GTGCGTTGAGCGCATGACAGGGCACCAAGAGTGGTCAACGATGTGAATATTCGCCGGA 1145  
Db 2159 GTGCGTTGAGCGCATGACAGGGCACCAAGAGTGGTCAACGATGTGAATATTCGCCGGA 2218

QY 1146 TGTAAGCTTAATTGCGTCTGCTTCATTTGACAAGTCACTGCGCTGTGTGGCGAGCCAGCGA 1205  
Db 2219 TGTAAGCTTAATTGCGTCTGCTTCATTTGACAAGTCACTGCGCTGTGTGGCGAGCCAGCGA 2278

QY 1206 TGGTCAGTACATGGCCACCTTCGGGGTTCATGTGCAGGCTGTTTACAGGTTGCCCTGGTC 1265



Db 2279 TGGTCAGTACATGGCCACCTTCCGGGTCATGTGAGGCTGTTTACACGGTTGCCCTGGTC 2338  
QY 1266 CGCGGACTCCCGCTTGATGTTTCCGGCAGCAAGAACTCAACTCTAAA----- 1313  
Db 2339 CGCGGACTCCCGCTTGATGTTTCCGGCAGCAAGAACTCAACTCTAAAAGGTTAGCACGT 2398  
QY 1314 -----AGTATGGAGTGTGCAGA 1330  
Db 2399 AGTTTCAGTTTCTAGCGCTAGCTAATACATATCCTCTTGCTTAGTATGGAGTGTGCAGA 2458  
QY 1331 CGAAGAAACTGGCACAGGAGCTGCCCTGGACATCGCGATGAGGTGTTCCGAGTGGACTGGG 1390  
Db 2459 CGAAGAAACTGGCACAGGAGCTGCCCTGGACATCGCGATGAGGTGTTCCGAGTGGACTGGG 2518  
QY 1391 CGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAAGTTATAAAGCT----- 1442  
Db 2519 CGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAAGTTATAAAGCTTT 2578  
QY 1443 -----ATGGGCTTA 1451  
Db 2579 GAGTAACGAGGTTTCTATAACAATATACAGACTCTAATGTATCTTTCAGATGGGCTTA 2638  
QY 1452 TTAACAAATCATTAACCTGTACACGGTAAGAAAATACCTAGGAATAAAGTAAAAACGTCCT 1511  
Db 2639 TTAACAAATCATTAACCTGTACACGGTAAGAAAATACCTAGGAATAAAGTAAAAACGTCCT 2698  
QY 1512 GAGTAAA 1518  
Db 2699 GAGTAAA 2705

RESULT 4  
ABL05232  
ID ABL05232 standard; cDNA; 6223 BP.  
XX  
AC ABL05232;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10178.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.

WO200171042-A2.  
27-SEP-2001.  
23-MAR-2001; 2001WO-US09231.  
23-MAR-2000; 2000US-191637P.  
11-JUL-2000; 2000US-0614150.  
(PEKE ) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;  
WPI; 2001-656860/75.  
P-PSDB; ABB61129.  
New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.  
Claim 1; SEQ ID NO 10178; 21pp + Sequence Listing; English.  
The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 6223 BP; 1458 A; 1588 C; 1574 G; 1603 T; 0 other;  
Query Match 38.7%; Score 601.2; DB 23; Length 6223;  
Best Local Similarity 99.0%; Pred. No. 1.8e-125;  
Matches 616; Conservative 0; Mismatches 3; Indels 3; Gaps 1;  
QY 7 CAAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCG 66  
Db 5602 CAGAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCG 5661  
QY 67 CTCGTTT--ACACGGGCGAGGAGCGCGGCCCAATCGACCTGCCGGCAGGAATCACT 123  
Db 5662 CTCGTTTCGGACACGGGCGAGGAGCGCGGCCCAATCGACCTGCCGGCAGGAATCACT 5721  
QY 124 ACCCAGCAATGGGACTGATTTGCAACCGCGTGTGAAAAACGAGGAAGCCACTCCATAT 183  
Db 5722 ACCCAGCAATGGGCTCTGATTTGCAACCGCGTGTGAAAAACGAGGAAGCCACTCCATAT 5781  
QY 184 TTGTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGACACGTTGGACTTGGCG 243  
Db 5782 TTGTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGACACGTTGGACTTGGCG 5841  
QY 244 TCAGTGGACACCGAAAAACGTGATCGATATTTGTATCAGCCACACAGCGCGGTTTCAAAGTG 303  
Db 5842 TCAGTGGACACCGAAAAACGTGATCGATATTTGTATCAGCCACACAGTGGGATTG 5901  
QY 304 CGCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 363  
Db 5902 CGCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 5961  
QY 364 AATTTACGCCCGGATGGTGTCTCATCTGCCAGTGGAAAGTGGCGACACAGTGGGATTG 423  
Db 5962 AATTTACGCCCGGATGGTGTCTCATCTGCCAGTGGAAAGTGGCGACACAGTGGGATTG 6021  
QY 424 TGGGATCTTAACACAGAGACACCGCACTTCACTGACAGGTCTAAGCAGTGGGTTCTG 483  
Db 6022 TGGGATCTTAACACAGAGACACCGCACTTCACTGACAGGTCTAAGCAGTGGGTTCTG 6081  
QY 484 TGCATATCCTGGGCTCCGGATGGCAACGGTTGGCAGCGGTTGCAAGCGGGCTCTATA 543  
Db 6082 TGCATATCCTGGGCTCCGGATGGCAACGGTTGGCAGCGGTTGCAAGCGGGCTCTATA 6141  
QY 544 ATCATCTGGGACCCCGGAGACGGGTGAGCAGAAGGGGGACCCCTTGAGTGGGCACAGAAA 603  
Db 6142 ATCATCTGGGACCCCGGAGACGGGTGAGCAGAAGGGGGACCCCTTGAGTGGGCACAGAAA 6201  
QY 604 CACATCAACTGCCTCGCCTGGG 625  
Db 6202 CACATCAACTGCCTCGCCTGGG 6223

RESULT 5  
ABL05214  
ID ABL05214 standard; cDNA; 4141 BP.  
XX  
AC ABL05214;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10124.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
OS Drosophila melanogaster.

PN WO200171042-A2.  
XX 27-SEP-2001.  
PD  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
DR P-PSDB; ABB61111.  
DR  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions .  
XX  
PS Claim 1; SEQ ID NO 10124; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pot\_sequences.  
XX  
SQ Sequence 4141 BP; 982 A; 1023 C; 1048 G; 1088 T; 0 other;

Query Match 37.8%; Score 588.2; DB 23; Length 4141;  
Best Local Similarity 81.9%; Pred. No. 1.4e-122;  
Matches 792; Conservative 0; Mismatches 3; Indels 172; Gaps 3;

QY 724 AACATTGCCGGACACACAAATGCTGTGACACAGCTGAGATGGGGTGGAGCGGCCTTAAT 783  
Db 1 AACATTGCCGGACACACAAATGCTGTGACACAGCTGAGATGGGGTGGAGCGGCCTTAAT 60  
QY 784 TATACATCCTCCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTTGTC 843  
Db 61 TATACATCCTCCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTTGTC 120  
QY 844 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAACAAACATTCGCTGAGCACCGATTACGTC 903  
Db 121 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAACAAACATTCGCTGAGCACCGATTACGTC 180  
QY 904 CTGCGCACCTGGTCCATTCCATCCGGTGAAGGATCGCTCAAGAGCCACCTCAGTTTGAGC 963  
Db 181 CTGCGCACCTGGTCCATTCCATCCGGTGAAGGATCGCTCAAGAGCCACCTCAGTTTGAGC 240  
QY 964 A-----C 965  
Db 241 AGTAAGAATTCATATGCTCTATATATGCTGTATATCTTAAAGAGATGTTTTACAGC 300  
QY 966 TGAGGAATTCAGGAATCTGCCTTGAAGCGCTACCGCCCGTGTGCCCTGACGAGTGGA 1025  
Db 301 TGAGGAATTCAGGAATCTGCCTTGAAGCGCTACCGCCCGTGTGCCCTGACGAGTGGA 360  
QY 1026 GTCGCTGGTTTCCTGTCGGATGACAAACCCCTCTATCTGTGGCGGAACCAACAGAACAA 1085  
Db 361 GTCGCTGGTTTCCTGTCGGATGACAAACCCCTCTATCTGTGGCGGAACCAACAGAACAA 420  
QY 1086 GTGCGTTGAGCGCATGACAGGGCACCAGAACGTTGTCACGATGTGAATATTCGCCGGA 1145  
Db 421 GTGCGTTGAGCGCATGACAGGGCACCAGAACGTTGTCACGATGTGAATATTCGCCGGA 480  
QY 1146 TGTAAGCTAATTGCGTCTGCTTTCATTGACAAAGTCAGTCGCTGTGGCGAGCCAGCGA 1205

Db 481 TGTAAAGCTAATTGCGTCTGCTTTCATTGACAAGTCAGTGCCTCTGTGGCGAGCCAGCGA 540  
QY 1206 TGGTCAGTACATGCGCCACCTTCCGGGGTCATGTGCAGGCTGTTTACACGTTGCCCTGGTC 1265  
Db 541 TGGTCAGTACATGCGCCACCTTCCGGGGTCATGTGCAGGCTGTTTACACGTTGCCCTGGTC 600  
QY 1266 CGCGGACTCCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAAAGGTTAGCACGT 1313  
Db 601 CGCGGACTCCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAAAGGTTAGCACGT 660  
QY 1314 -----AGTATGGAGTGTGCAGA 1330  
Db 661 AGTTTCAGTTTCTAGCGCTAGCTAATACATATCCTCTTGCTTAGTATGGAGTGTGCAGA 720  
QY 1331 CGAAGAAACTGGCACAGGAGCTGCCCTGGACATGCCGATGAGGTGTTCCGAGTGGACTGGG 1390  
Db 721 CGAAGAAACTGGCACAGGAGCTGCCCTGGACATGCCGATGAGGTGTTCCGAGTGGACTGGG 780  
QY 1391 CGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAGTTATAAAGCT----- 1442  
Db 781 CGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAGTTATAAAGCTGTAAGTTT 840  
QY 1443 -----ATGGGCTTA 1451  
Db 841 GAGTAACGAGGTTTCTATAACAATATACAGACTCTAATTGTATCTTTCAGATGGGCTTA 900  
QY 1452 TTAACAAATCATTAACTTGACACGCGTAAGAAATACTTAGGAATAAAAGTAAACGTCCT 1511  
Db 901 TTAACAAATCATTAACTTGACACGCGTAAGAAATACTTAGGAATAAAAGTAAACGTCCT 960  
QY 1512 GAGTAAA 1518  
Db 961 GAGTAAA 967

RESULT 6  
AAF85075  
ID AAF85075 standard; DNA; 1859 BP.  
XX  
AC AAF85075;  
XX  
DT 09-JUL-2001 (first entry)  
XX  
DE Nucleotide sequence of a human Tzap gene cDNA clone.  
KW Tzap; T cell activation; immune response; transplant rejection;  
KW bone marrow transplantation; rheumatoid arthritis; lupus erythematosus;  
KW multiple sclerosis; encephalitis; vasculitis; diabetes mellitus;  
KW pancreatitis; gastritis; thyroiditis; malignant disorder; asthma;  
KW lepraematosis; gastritis; skin tumour; adrenal tumour; lung tumour;  
KW wound healing; growth disorder; inflammatory disease; infectious disease;  
KW ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 20..1477  
FT /\*tag= a  
FT /product= "Tzap"  
FT /transl\_except= "(pos: 1154..1156, aa: Xaa)"  
FT /note= "Xaa represents an unspecified residue"  
XX  
PN WO200132614-A2.  
XX  
PD 10-MAY-2001.  
XX  
PF 30-OCT-2000; 2000WO-EP10670.  
XX  
PR 01-NOV-1999; 99US-0162675.  
PR 25-FEB-2000; 2000US-0185016.  
XX  
PA (UTKU/) UTKU N.









Db 1460 CAAATGCCCTCCGGATATGG 1478  
RESULT 8  
ID AAF58316  
XX AAF58316 standard; cDNA; 1854 BP.  
AC AAF58316;  
XX 19-APR-2001 (first entry)  
XX Human GTP-binding associated protein #16 coding sequence.  
DE Human; guanosine triphosphate binding associated protein; GTP; GBAP;  
KW inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma;  
KW autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;  
KW osteoporosis; psoriasis; ss.  
XX Homo sapiens.  
OS WO200105970-A2.  
XX 25-JAN-2001.  
XX 19-JUL-2000; 2000WO-US19698.  
XX 19-JUL-1999; 99US-0144595.  
XX 23-AUG-1999; 99US-0150460.  
XX 15-OCT-1999; 99US-0159849.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Yue H, Tang YT, Bandman O, Hillman JL, Lal P, Au-Young J;  
PI Reddy R, Yang J, Baughn MR, Lu DAM, Azimzai Y, Patterson C;  
XX WPI; 2001-091972/10.  
XX P-PSDB; AAB68516.  
XX New guanosine triphosphate-binding associated proteins (GBAP) and their  
PT encoding nucleic acids, useful for treating and/or diagnosing diseases  
PT associated with GBAP expression, such as cancer, diabetes and asthma -  
XX Claim 5; Page 197; 233pp; English.  
XX The present invention relates to novel human guanosine triphosphate  
CC (GTP)-binding associated proteins (GBAPs; AAB68501-AAB68566) and their  
CC coding sequences (AAF58301-AAF58366). The proteins and coding sequences  
CC of the present invention are useful for treating a variety of disorders  
CC including inflammation, AIDS, Addison's disease, anaemia,  
CC arteriosclerosis, asthma, autoimmune disorders, Grave's disease,  
CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and  
XX psoriasis.  
SQ Sequence 1854 BP; 389 A; 548 C; 563 G; 354 T; 0 other;  
Query Match 30.7%; Score 477.4; DB 22; Length 1854;  
Best Local Similarity 59.8%; Pred. No. 9.2e-98;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;  
QY 54 GATACAGGCGCCCTCGTTTACACGGCGGAGGAGCCGCGCCCAATCGACCTGCCGCG 113  
Db 69 GCTAGTGCAGTTCAGGATGAGGGCGGCAGCTGCTGGGTTCCCGTTCCAGCTGCCCT 128  
QY 114 AGGAATCACTACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGC 173  
Db 129 GGACATCACCCCGACAGGCTGCAGCTCGTGTGCAACGCGCTACTGGCCCGAGGAGATCC 188  
QY 174 CACTCCATATTTGTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGCTT 233  
Db 189 CCTGCCACTGGCTTCTCTTGTCCAGGATGCTGAGATCGTCTCCTCACTGGGGAAGACGTT 248  
QY 234 GGACTTGGCGTCACTGGACACCCGAAACAGTGCATATTTGTATATCAGCCACAGGCGGT 293

Db 249 GGAGTCCAGGCGAGTGGAGACAGAGAAGGTCTCTAGACATCATCTACCAGCCACAGGCTAT 308  
QY 294 TTTCAAAGTGCGCCCACTGACAAAGATGCACAGAGTCCATGCCGGGACACGCGAGGCTGT 353  
Db 309 CTTTCAGAGTCCGGGCTGTGACTCGTGCACACAGTCTCTTGGAGGCTCACAGTGAAGCAGT 368  
QY 354 GGTTCGCTGAATTTACAGCCCGGATGGTGTCTCATCTCGCCAGTGGGAAGTGGCGACACAC 413  
Db 369 CATTTCTGTGGCCTTCAGCCCTACGGGAAAGTACTGTGCCAGTGGCTCTGGAGACACCA 428  
QY 414 AGTGGATGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCA 473  
Db 429 CGTGCCTTCTGGGATCTCAGCACAGAGACACCATTTTCATGTCGAAGGACACAGACA 488  
QY 474 GTGGTTCGTGCGTATCTCTGGGCTCCGGATGGCAAAACGGTTGCCAGCGGTTGCAAGC 533  
Db 489 CTGGTCCCTAGTATATCTCTGGTCTCCAGATGGCAAGAAAGCTGGCCTCAGGCTGCNAGAA 548  
QY 534 GGGCTCTATAATCATCTGGGACCCGGAGACGGGTACAGAGAGGGGCGACCTTGTAGTGG 593  
Db 549 TGGCCAGATCTCCTCTGGGACCCCAAGCACAGGAAGCAGGTGGGAGGACCTCGCTGG 608  
QY 594 GCACAAGAAACACATCAACTGCCCTGCCCTGGGACCGGTATCATCGCGATCCGAGTGCAG 653  
Db 609 CCACAGCAAGTGGATCACAGGCTGTGAGCTGGGAGCCCTCCATCGGAACCTGTAGTGCCG 668  
QY 654 GAAACTTGTCTCCGCCAGTGGAGACGGGGACTGCGCGGATTTGGACGTAATAATTGGGCCA 713  
Db 669 CTATGTGCCAGCAGCTCCCAAGGATGGCAGTGTGCGGATCTGGGACACAACTGCAGGCCG 728  
QY 714 GTGCCTTATGAACATTCGCCGACACACAAAATGCTGTGACAGAGTGGGTGGGAGC 773  
Db 729 CTGTAGCGCATCTCACCGGGCACACCCAGTCGCTCACCTGTCTCCGTGGGAGGGGA 788  
QY 774 GGGCCTTATTTATACATCTCCAAAGATCGCACAGTGAAGATGGCGGAGAGCTGATGG 833  
Db 789 CGGCTTCTCTACTCTGCTCCAGGACCGCACCATCAAGTCTGGAGAGCTCATGACGG 848  
QY 834 AATCTTGTCCCGGACGTTCTCTGSCCAAGCTCACTGGGTAAACAACATTTGGCTGAGCAC 893  
Db 849 TGTGCTGTGCGGACTCTGCAAGSCCAGCGCCACTGGGTGAACACCATGGCCCTCAGCAC 908  
QY 894 CGATTACGTCCTGCGCACTGGTCCATTCATCCGCTGAAGGATCGCTCCAAGAGCC--A 950  
Db 909 TGACTATGCCCTGCGCACTGGGGCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCAAGA 968  
QY 951 CCTCAGTTTGAGCACTGAGGAATTCAGGAATCTGCTTGAAGCGCTACCAGGCCGTGTG 1010  
Db 969 CCTCCAAGGATCCTTGCAGGAGTTGAAGGAGAGGCTCTGAGCGGATACAAACCTCGTGG 1028  
QY 1011 CCCTGACGAGGTGGAGTGGTTCCTGTTTCCTGTTGCGATGACAAACCTCTATCTGTGG-- 1068  
Db 1029 GGGCCAGGTCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACCTTATTCCTGTGGTC 1088  
QY 1069 -CGGAACACACAGAACAAAGTGGTTCGCTGAGCGCATGACAGGGCACAGAACGTTGTCACGA 1127  
Db 1089 CCCAGCAGAGGACAAAAGCCTCTCTACTCGGATGACAGGACACCAAGCTCTCATCAACCA 1148  
QY 1128 TGTGAAATATTCGCGGATGTAAAGCTAAATTCGCTGTCTTCAATTTGACAAAGTCAAGTGG 1187  
Db 1149 GGTGCTCTCTCTCTCTCCGACTCCCGCATCGTGGTGTAGTGCCTCTCTTGAACAGTCCATCA 1208  
QY 1188 TCTGTGGCAGCCAGCGATGGTCACTGATGCGCCACTTCCGGGTGATGTGGAGGCTGT 1247  
Db 1209 GCTGTGGGATGGCAGGACGGGCAAGTACCTGGCTTCCCTACGGCGGCCACGTTGGCTGCCGT 1268  
QY 1248 TTACAGGTTGCCCTGGTCCCGGACTCCCGCTTGTATGTTTCCGGCAGCAAGACTCAAC 1307  
Db 1269 GTACCAGATGCGTGGTTCAGCTGACAGTGGCTCCCTGCTGAGCGGCGAGCTGACAGCAC 1328  
QY 1308 TCTAAAAGTATGAGTGTGCAGACGACGAAACTGSCACAGGAGCTGCCTGGACATGCGGA 1367

Db 1329 ACTGAAGGTGTGGGATGTGAAGCCAGAAAGCTGGCCATGGACCTGCCCGCCACGCGGA 1388  
QY 1368 TGAGGTGTTTCGGAGTGGACTGGGCGCCCGATGGCTCTAGAGTTGCCCTGTGGTGGCAAGGA 1427  
Db 1389 TGAGGTATATGCTGTTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA 1448  
QY 1428 CAAAGTTATAAAGCTATGG 1446  
Db 1449 CAAATGCCCTCCGGATATGG 1467

RESULT 9  
ABN59719  
ID ABN59719 standard; cDNA; 2638 BP.  
XX ABN59719;  
AC ABN59719;  
XX  
DT 28-JUN-2002 (first entry)  
XX Novel human coding sequence SEQ ID NO: 130.  
DE Human; antianaemic; vulnery; antiinflammatory; immunomodulator;  
XX antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;  
KW neuroprotective; antiparkinsonian; protein therapy; Est;  
KW expressed sequence tag; gene; ss.  
XX

OS Homo sapiens.  
XX WO200222660-A2.  
PN 21-MAR-2002.  
XX 10-SEP-2001; 2001WO-US26015.  
PF 11-SEP-2000; 2000US-0659671.  
PR (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX WPI; 2002-292408/33.  
DR P-PSDB; ABB97306.

PT An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis -  
PS Claim 1; SEQ ID NO 130; 509pp; English.

CC The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat  
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions  
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a coding sequence of the  
CC invention.

XX Sequence 2638 BP; 566 A; 767 C; 764 G; 541 T; 0 other;  
SQ  
Query Match 30.7%; Score 477.4; DB 24; Length 2638;  
Best Local Similarity 59.8%; Pred. NO.1e-97;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 54 GATACAGGCGCCCTCGTTTACAGGGCGGAGGAAGCCGGCCCGCAATCGACCTGCCGGC 113  
Db 132 GCTAGTGCAGTTCAGGATGAGGGCGGCGAGCTGCTGGGTTCCTCGACGTGCCCGT 191  
QY 114 AGGAATCACTACCCAGCAATTGGGACTGATTTGCAAGCGCTGCTGAAAAACGAGGAAGC 173

Db 192 GGACATCACCCCGACAGGCTGCAGCTCGTGTGCAACGGCTACTGGCCACGAGGATCC 251  
QY 174 CACTCCATATTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCCTGGAGGACAGCTT 233  
Db 252 CCTGCCACTGGCTTCTTTGTCACGATGCTGAGATCGTCTCCTCACTGGGGAAGAGCTT 311  
QY 234 GGACTTGGCGTCACTGGACACCGAAACGTGATCGATATGTTGATCAGCCACAGCGGT 293  
Db 312 GGAGTCCAGGAGTGGAGACAGAGAGAGTCTCTAGACATCATCTACGAGCCACAGGTAT 371  
QY 294 TTTCAAAGTGGCCCGAGTGACAAAGATGCACGAGTTCATGCCGCGGAGACGCCGAGCTGT 353  
Db 372 CTTACAGAGTCCGGGCTGTGACTGCTGCACAGTCTCCTGGAGGGTACAGTGAAGCAGT 431  
QY 354 GGTTCGCTGAATTTCAAGCCCGATGGTGTCTCATCTGCCAGTGGAGTGGCGACACCAC 413  
Db 432 CATTTCTGTGGCTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACCAC 491  
QY 414 AGTGGATTTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCA 473  
Db 492 CGTGGCTTCTGGGATCTCAGACAGAGACACCACTTTCATATGCAAGGGACAGACA 551  
QY 474 GTGGTTCGTGCTGATCTCTGGGCTCCGGATGGCAAGCGTTGGCCAGCGTTGCAAAAGC 533  
Db 552 CTGGGTCTTAGTATATCTGTCTCCAGATGGCAAGAGTGGCTCAGGCTGCAAGAA 611  
QY 534 GGGCTCTATATCATCTGGGACCCGGAGAGCGGTCAGCAGAAGGGCGACCCCTTGTAGTGG 593  
Db 612 TGGCCAGATTCTCTCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGGAGGACCCCTCGCTGG 671  
QY 594 GCACAAGAAACACATCAACTGCCTCGCTGGGAAACCGTATCATCGCGATCCGGAGTGCAG 653  
Db 672 CCACAGCAAGTGGATCACAGGCTGAGCTGGGAGGCCCTCCATGCGAACCCTGAGTGGCG 731  
QY 654 GAAACTTGTTCGCCAGTGGAGAGCGGGGACTGCCGATTTGGACGTAAATTTGGGCCCA 713  
Db 732 CTATGTGGCCAGAGCTCCAAGGATGGCAGTGTGGGACACAACTGCAAGGCCG 791  
QY 714 GTGCTTATGAACATTTCCGGACACACAAATGCTGTGACAGCAGTGAAGTGGGTGGAGC 773  
Db 792 CTGTGAGCGCATCTCTCACCGGACACCCAGTGGTCACTGTCTCCGGTGGGAGGGA 851  
QY 774 GGGCTTATTATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG 833  
Db 852 CGGGCTTCTCTACTCTGCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG 911  
QY 834 AATCTTGTCCGGACGTTCTCTGGCCAAAGTCACTGGGTAAACAAATGCGCTGAGCAC 893  
Db 912 TGTGCTGTCCGGACTCTGCAAGGCCACGCGCCACTGGGTGAACACCATGGCCCTCAGCAC 971  
QY 894 CGATTACGCTCTGCGCACCTGGTCCATTCCATCCGGTGAAGGATCGTCCCAAGAGCC---A 950  
Db 972 TGACTATGCCCTGCGCACCTGGGGCCCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCAAGA 1031  
QY 951 CCTCAGTTTGAGCACTGAGGAATTCAGGAATCTGCCCTGAAGCGCTACCAGGCGGTGTG 1010  
Db 1032 CCTCAAAGGATCTTGCAGGAGTTGAAGGAGAGGGCTCTGAGCCGATACAACTCGTGGC 1091  
QY 1011 CCCTGACGAGTGGAGTGGTTCCTGTTTCGATGACAAACACACCTCTATCTGTGG-- 1068  
Db 1092 GGGCCAGGTTCCAGAGAGGCTGGTGTCTGGCTCCGAGCACTTACCTTATTCCTGTGGTC 1151  
QY 1069 -CGAACAACCCAGAACAAAGTGGCTTGAGCGCATGACAGGGCACCCAGAACGTTGGTCAACGA 1127  
Db 1152 CCACAGAGAGGACAAAGCCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA 1211  
QY 1128 TGTGAAATATTCGCCCGGATGTAAGCTAATTCGCTGCTTTCATTCACAAGTCAAGTGG 1187  
Db 1212 GGTGCTCTCTCTCTGACTGCCGATCGTGGCTAGTGGCTCTCTTGACAAGTCCATCAA 1271  
QY 1188 TCTGTGGCGAGCCAGCATGGTTCAGTACATGGCCACCTCCGGGGTCATGTGAGGCTGT 1247  
Db 1272 GCTGTGGGATGGCAGGACGGGCAAGTACCTGGCTTCCTTACGGCGCCACGTGGTGGCT 1331



QY	951	CCTCAGTTTGAGCACTGAGGAATTGCAGGAATCTGCCTTTGAAGCGCTACCGAGGCCGTGTG	1010	PR	28-APR-1999;	99US-0131449.
Db	1039	CTTCCAAGGATCCTTGCAGAGATTGAAGAGAGAGGGCTCTGAGCGGATACAACTCGTGCG	1098	PR	30-APR-1999;	99US-0132048.
QY	1011	CCCTGACGAGGTGGAGTCGCTGGTTTCCTGTTCCGATGACACACCCCTCTATCTGTGG--	1068	PR	04-MAY-1999;	99US-0132407.
Db	1099	GGCCAGGTCACAGAGAGGCTGGTGTCTGGCTCCGACGACTTACCTTATCTCTGTGGTC	1158	PR	05-MAY-1999;	99US-0132485.
QY	1069	-CGGAACAACCAAGTGCCTTGAGCGCATGACAGGGCACAGAACGTGGTCAACGA	1127	PR	06-MAY-1999;	99US-0132486.
Db	1159	CCGACGAGAGACAAAAGCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA	1218	PR	07-MAY-1999;	99US-0132487.
QY	1128	TGTGAATATTCGCCGGATGTAAAGCTTAATTCGCTCTGCTTCATTTGACAAGTCAGTGG	1187	PR	11-MAY-1999;	99US-0132863.
Db	1219	GGTGCTCTTCTCTCCTGACTCCCGCATCGTGGCTAGTGCCTCTTGACAAGTCCATCAA	1278	PR	14-MAY-1999;	99US-0134256.
QY	1188	TCGTGGCGGACCGGATGGTCAGTACATGGCCACCTTCCGGGTCATGTGCAGGCTGT	1247	PR	14-MAY-1999;	99US-0134218.
Db	1279	GCTGTGGGATGCGAGGACGGCAAGTACCTGGCTTCCCTACGGGCCACGTGGCTGCCGT	1338	PR	14-MAY-1999;	99US-0134219.
QY	1248	TTACACGGTTGCCTGGTCCCGGGACTCCCGCTTGATTTTCCGGCAGCAAGACTCAAC	1307	PR	14-MAY-1999;	99US-0134221.
Db	1339	GTACCAGATTGCGTGGTCAGCTGACAGTCCGCTCCTGGTCAGCGGCAGCACTGACAGCAC	1398	PR	18-MAY-1999;	99US-0134370.
QY	1308	TCFAAAGTATGGAGTGTGCAGACGAAAGAACTGGCAGGAGTGCCTGGACATGCCGA	1367	PR	19-MAY-1999;	99US-0134941.
Db	1399	ACTGAAGGTGGGATGTGAAGGCCCAAGAGCTGGCCATGGACTGCCGCCACCGCGGA	1458	PR	20-MAY-1999;	99US-0135124.
QY	1368	TGAGGTGTTCCGAGTGGACTGGGCGCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGA	1427	PR	21-MAY-1999;	99US-0135353.
Db	1459	TGAGGTATATGCTGTTGACTGGAGTCCAGATGGCCAGAGATGGCAAGTGGTGGGAAGGA	1518	PR	24-MAY-1999;	99US-0135629.
QY	1428	CAAAGTTATAAAGCTATGG	1446	PR	25-MAY-1999;	99US-0136021.
Db	1519	CAAATGCCTCCGGATATGG	1537	PR	27-MAY-1999;	99US-0136392.
RESULT 11						
ID	AAC48725	standard; DNA; 1726 BP.				
XX	XX					
AC	AAC48725;					
XX	XX					
DT	18-OCT-2000	(first entry)				
XX	XX					
DE	Arabidopsis thaliana	DNA fragment SEQ ID NO: 58533.				
XX	XX					
KW	Hybridisation assay; genetic mapping; gene expression control;					
KW	protein identification; signal transduction pathway;					
KW	metabolic pathway; promoter; termination sequence; ss.					
XX	XX					
OS	Arabidopsis thaliana.					
XX	XX					
PN	EP1033405-A2.					
PD	06-SEP-2000.					
XX	XX					
PF	25-FEB-2000; 2000EP-0301439.					
XX	XX					
PR	25-FEB-1999;	99US-0121825.				
PR	05-MAR-1999;	99US-0123180.				
PR	09-MAR-1999;	99US-0123548.				
PR	23-MAR-1999;	99US-0125788.				
PR	25-MAR-1999;	99US-0126264.				
PR	29-MAR-1999;	99US-0126785.				
PR	01-APR-1999;	99US-0127462.				
PR	06-APR-1999;	99US-0128234.				
PR	16-APR-1999;	99US-0128714.				
PR	19-APR-1999;	99US-0130077.				
PR	21-APR-1999;	99US-0130449.				
PR	23-APR-1999;	99US-0130510.				
PR	23-APR-1999;	99US-0130891.				



PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	21-JUL-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	26-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	26-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	26-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161992.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR	29-OCT-1999;	99US-0162142.
PR	04-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149722.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149902.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
PR	13-SEP-1999;	99US-0153758.			
PR	15-SEP-1999;	99US-0154018.			
PR	16-SEP-1999;	99US-0154039.			
PR	20-SEP-1999;	99US-0154779.			
PR	22-SEP-1999;	99US-0155139.			
PR	23-SEP-1999;	99US-0155486.			
PR	24-SEP-1999;	99US-0155659.			
PR	28-SEP-1999;	99US-0156458.			
PR	29-SEP-1999;	99US-0156596.			
PR	04-OCT-1999;	99US-0157117.			
PR	05-OCT-1999;	99US-0157753.			
PR	06-OCT-1999;	99US-0157865.			
PR	07-OCT-1999;	99US-0158029.			
PR	08-OCT-1999;	99US-0158232.			
PR	12-OCT-1999;	99US-0158369.			
PR	13-OCT-1999;	99US-0159293.			
PR	13-OCT-1999;	99US-0159294.			
PR	13-OCT-1999;	99US-0159295.			
PR	14-OCT-1999;	99US-0159329.			
PR	14-OCT-1999;	99US-0159330.			
PR	14-OCT-1999;	99US-0159331.			
PR	14-OCT-1999;	99US-0159637.			
PR	14-OCT-1999;	99US-0159638.			

Query Match	21.4%;	Score 333.4;	DB 21;	Length 1726;
Best Local Similarity	54.4%;	Pred. No. 2.2e-65;		
Matches 723;	Conservative	0;	Mismatches 591;	Indels 15; Gaps 2;

QY	129	GCAATTGGGACTGATTGTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATATTGTT	188
Db	163	GCAACTTACTCAGCTCGTCAATAGCTTCTTGTGACAAACGAGGAGATGTTACCTTACAGTTT	222
QY	189	TTTCGTGGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTTGGACTTGGCGTCAGT	248
Db	223	CTATGTATCAGACGAGAGCTTCTTGTACCACTTGGAACTACTTGGAGAAAAACAAGT	282
QY	249	GGACACCGAAACGTCATCGATATTGTGTATCAGCCACAGGCGGTTTTCAAAGTGGCCCC	308
Db	283	GTCGTGGAGAAGTTTGTGACGATTGTTTATCAACAACAAGCTGTTTTCGAATTCGTCC	342
QY	309	AGTGACAAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTGAATTT	368
Db	343	TGTTAACCGTTGCTCAGACAGACAATTGCTGGTCACGCGGAGCTGTCTTTGRTTTCGTT	402
QY	369	CAGCCCGGATGGTCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCAGATTGTGGGA	428
Db	403	TAGTCTGTATGGAAGCAATTAGCTAGTGGTTTCAGGTGATACTACTGTCTCAGGCTTTGGGA	462
QY	429	TCTTAACACAGAGACACCGCACTTCCATGCCGGGACACAGCTGCACAGTGCATGAAGCGGTCTGTGCGT	488
Db	463	TCTCTACACTGAAACTCCATTGTTTACTTGCAAAGCCACAGAATTTGGTTCTCTCACAGT	522
QY	489	ATCCTGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAGCGGGCTCTATAATCAT	548
Db	523	TGCGTGGTCTCCTGATGGTAAAGCATCTTGTGAGTGTAGTAAATCCGGTGAATCTGTTG	582
QY	549	CTGGACCCGGAGACGGGTGACGAGAAGGGGGACCCCTTGAGTGGGCACAGAACAACAT	608
Db	583	TTGGAATCCAAAGAAGGAGAACTAGAAGGCAGCCCACTTACGGGTCACAGAAGATGGAT	642
QY	609	CAACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACCTTCCTCCGC	668
Db	643	TACTGGTATCTCGTGGGAACCCAGTCCACCTTAGTTCTCCATGCCGTGATTTGTGACTTC	702
QY	669	CAGTGGAGACGGGACTGCCGGATTGCGGACGTAATAATTGGGCAGTGCCTTATGAACAT	728
Db	703	TAGTAAAGATGGGGATGCAAGGATTGGGATATTACACTGAAAAAATCTATTATTGTCT	762
QY	729	TGCCGGACACAAAATGCTGTGACAGAGTGAGATGGGGTGGACGGGCTTATTATTATAC	788
Db	763	CAGTGGCACACACTTGTGTGACTTGTGTCAAAATGGGGCGGAGATGGAATTTATTATAC	822
QY	789	ATCCTCCAAAGATCGCACAGTGAAGATGTGGCGGACGAGCTGATGGAATCTTGTGCCGGAC	848
Db	823	AGGTCGCAAGATTGTACGATTAAAGATGTGGGAGACTACTCAGSGGGAAGCTTATTCGTGA	882

QY	849	GTTCTCTGGCCAAAGCTCACTGGGGTAAACAACATTTGCCGTGAGCACCGATTACGTCCTGGC	908	PR	01-APR-1999;	99US-0127462.
Db	883	ATTGAAGGGCATGGGCATTGGATTAACTCCCTTGGCTGAGCACAGATAATGTCCTCG	942	PR	06-APR-1999;	99US-0128234.
QY	909	CACITGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGACCACTGA	968	PR	08-APR-1999;	99US-0128714.
Db	943	AACAGGAGCTTTTGACC-----ACACTGGAAGACAAATATCCTCCRAATGAAGAAAGCAA	997	PR	16-APR-1999;	99US-0129845.
QY	969	GGAAATGCAGGAATCTGCCCTTGAAGCGCTACCAGGCGGTGTGCCCTGACGAGGTGGAGTC	1028	PR	19-APR-1999;	99US-0130077.
Db	998	AAGGCGCTCGAAAGATACAAACAACAAAGGGGATTCCCTGAAAGATTAGTCTCAGGT	1057	PR	21-APR-1999;	99US-0130449.
QY	1029	GCTGGTTTCCGTTCGGATGACAACACCCTCTATCTGTGGCGGAACAACCAACAAGTG	1088	PR	23-APR-1999;	99US-0130510.
Db	1058	TCGTGATGATTTCACATATGTTCCCTTTGGGAACCATCTGTAGCAAAACCACTAAA-----	1111	PR	23-APR-1999;	99US-0130891.
QY	1089	CGTTGAGCGCATGACAGGGCACAGAACCTGGTCAACGATGTGAATATTCGCCGGATGT	1148	PR	28-APR-1999;	99US-0131449.
Db	1112	----AAGCGCTTAACCGGTCAATCAACAGCTTGTAAATCATGTCTATTTCTCAGCTGATGG	1167	PR	30-APR-1999;	99US-0132048.
QY	1149	AAAGCTAATGCGTCTGCTTCATTTGACAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT	1208	PR	04-MAY-1999;	99US-0132484.
Db	1168	GAATGGATTGCAAGTGCATCGTTTCGATAAATCAGTAGGCTATGGAATGGTATCACAGG	1227	PR	05-MAY-1999;	99US-0132485.
QY	1209	TCAGTACATGGCCACCTTCGGGGTCAATGTGCAGGCTGTATTACACGGTTGCCGTCCGC	1268	PR	06-MAY-1999;	99US-0132486.
Db	1228	ACAATTTGTTACAGTTTTCGGGGGACATGTTGGACCTGTTATCAGGTCAGTTGGTCCGC	1287	PR	07-MAY-1999;	99US-0132863.
QY	1269	GGACTCCCGCTGATTGTTTCCGGGCAGCAAAAGACTCAACTCTAAAGTATGGAGTGTGCA	1328	PR	11-MAY-1999;	99US-0134256.
Db	1288	AGACAGTAGATTGCTTTTGAGTGGCAGTAAAGACTCTACTCTCAAGATATGGAAATTAG	1347	PR	14-MAY-1999;	99US-0134218.
QY	1329	GACGAAGAACTGGCACAGGAGTGCCCTGGACATCGGATGAGGTGTTTCGGAGTGGACTG	1388	PR	14-MAY-1999;	99US-0134219.
Db	1348	GACGAAAAAGTTAAAAACAAGATCTTCCTGGTCAATGCAGATGAGGTTTTCGGTGGATTG	1407	PR	14-MAY-1999;	99US-0134221.
QY	1389	GGCGCCCGATGGCTCTAGACTTGCCTGGTGGCAAGGACAAAGTTATAAAGCTATGGGC	1448	PR	14-MAY-1999;	99US-0134370.
Db	1408	GAGTCCAGATGGAGAGAAAGTAGTTTCTGGTGTAAAGATAGAGTGTGAAGCTATGGAA	1467	PR	18-MAY-1999;	99US-0134768.
QY	1449	TTATTAACA 1457		PR	19-MAY-1999;	99US-0134941.
Db	1468	GGGTTAAAA 1476		PR	20-MAY-1999;	99US-0135124.
RESULT 12						
ID	AAC39906 standard; DNA; 1643 BP.					
XX	AAC39906;					
AC	(first entry)					
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 26334.					
DT	Hybridisation assay; genetic mapping; gene expression control;					
XX	protein identification; signal transduction pathway;					
DE	metabolic pathway; promoter; termination sequence; ss.					
KW	Arabidopsis thaliana.					
KW	EPI033405-A2.					
XX	06-SEP-2000.					
PD	25-FEB-2000; 2000EP-0301439.					
XX	25-FEB-1999; 99US-0121825.					
PR	05-MAR-1999; 99US-0123180.					
PR	09-MAR-1999; 99US-0123548.					
PR	23-MAR-1999; 99US-0125788.					
PR	25-MAR-1999; 99US-0126264.					
PR	29-MAR-1999; 99US-0126785.					



QY	729	TGCCGGACACACAATGCTGTGACAGCAGTGGGTCGAGCGGGCTTATTATATAC	788
Db	763	CAGTGGGCACACACTGCTGTGACTTGTGCAAAATGGGCGGAGATGGAATATTATAC	822
QY	789	ATCCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGCCGGAC	848
Db	823	AGTTTCGAAGATTGACGATTAAAGATGTGGGAGACTACTCAGGGGAAGCTTATTCTGTA	882
QY	849	GTCTCTGCGCCCAAGTCACTGGSTAAACAACATTCGCTGAGCACCAGGATTACGTCCTGGG	908
Db	883	ATTGAAGGGCATGGGCATTGGATTAACCTCCCTGCGTTGAGCACAGAAATATGTTCTCG	942
QY	909	CACCTGTCCCATTCATCCCGTGAAGGATCGCTCCAAAGAGCCACCTCAGTTTGAGCACTGA	968
Db	943	AACAGGAGCTTTTGACC-----ACACTGGAAGACAATATCTCCAAATGAAGAAAGCAA	997
QY	969	GGAATTGCAGGAATCTGCCCTTGAGCGCTACACAGCCGCTGTGTCCTGACAGAGTGGAGTC	1028
Db	998	AAGCGCTCGAAGATACACAACAACAAGGGATTCCTGAAAGATTAGTCTCAGGT	1057
QY	1029	GCTGGTTCCCTGTCGGATGACACACACCTCTATCTGTGCGGGAACACACAGCAAGTG	1088
Db	1058	TCTGATGATTTCACTATGTTCTCTTGGGAACCATCTGTTAGCAACAACCTAAA-----	1111
QY	1089	CGTTGAGCGCATGACAGGGCACCAAGAACGTTGGTCAACGATGTGAAATATCGCCGGATGT	1148
Db	1112	----AAGCGCTTAACCGGTCAACAACAGCTTGTAAATCATGTCTATTCTCACCTGATGG	1167
QY	1149	AAAGCTAAATGGTCTGCTTCAATTGACAAGTCAGTGCCTGTGTGGCGAGCCAGCGATGG	1208
Db	1168	GAAATGGATTGCAAGTGCATCGTTCGATAAATCAGTTAGTTATGGAATGGTATCACAGG	1227
QY	1209	TCAGTACATGGCCACCTTCCGGSGTCATGTGCAGGCTGTTTACACGGTTCCTGGTCCGC	1268
Db	1228	ACAATTTGTACAGTTTCCGGGGCCATGTTGGACCTGTTTATCAGGTCAGTTGGTCCGC	1287
QY	1269	GGACTCCCGCTTGATGTTTCCGGCGAGCAAAAGACTCAACTCTAAAAGATGAGTGTGCA	1328
Db	1288	AGACGTAGATTGCTTTTGAGTGGCAGTAAAGACTCTACTCTCAAGATATGGAAATTAG	1347
QY	1329	GACGAAGAAACTGGCACAGGAGTGCCTGGACATGGGATGAGTGTTCGGAGTGGACTG	1388
Db	1348	GACGAAAAGTTAAAACAAGATCTTCTGTCATGCTGATGAGGTTTTGCGGTGGATTG	1407
QY	1389	GGCGCCGATGGCTCTAGAGTTGCCTCTGTGGCAAGGACAAAAGTTATAAAGCTATGGGC	1448
Db	1408	GAGTCCAGATGGAGAAAAGTAGTTTCTGTGTGTAAGATAGAGTGTGAAGCTATGGAA	1467
QY	1449	TTATTAAACA	1457
Db	1468	GGGTAAAA	1476
RESULT 13			
AAF85074			
ID	AAF85074	standard; DNA; 1170 BP.	
XX			
AC	AAF85074;		
XX			
DT	09-JUL-2001	(first entry)	
XX			
DE		Nucleotide sequence of human Tzap gene cDNA clone Tzap7/A.	
XX			
KW		Tzap; T cell activation; immune response; transplant rejection;	
KW		bone marrow transplantation; rheumatoid arthritis; lupus erythematosus;	
KW		multiple sclerosis; encephalitis; vasculitis; diabetes mellitus;	
KW		pancreatitis; gastritis; thyroiditis; malignant disorder; asthma;	
KW		lepramatosus; gastritis; skin tumour; adrenal tumour; lung tumour;	
KW		wound healing; growth disorder; inflammatory disease; infectious disease;	
OS		Homo sapiens.	
XX			



Db 361 TGGTGTCTGGTCCGACGACTTACCTTATTCCTGTGGTCCCCAGCAGAGACAAAAGC 420  
Qy 1088 GCCTTGAGCGCATGACAGGACACAGACGCTGGTCAACGATGGAATATTCGCCGGATG 1147  
Db 421 CTCTCACTCGGATGACAGACACCAAGCTCTCATCAACAGGCTCTCTCTCTGACT 480  
Qy 1148 TAAAGCTAATTCGGCTGCTTTCATTGACAACTCAGTCCGCTGTGGGAGCCAGCGATG 1207  
Db 481 CCGCATCGTGGCTAGTGCCTCTTGGACAAGTCCATCAAGCTGTGGGATGGCAGGACGG 540  
Qy 1208 GTCACTACATGCGCCACCTTCCGGGGTCATGTGCAGGCTGTTTACACGGTTCCTGGTCCG 1267  
Db 541 GCAAGTACCTGGCTTCCCTACGCGGCCACGTGGCTGCGGTACACAGATTGCGTGGTCAG 600  
Qy 1268 CGGACTCCCGCTTGATTGTTTCCGGGACCAAGACTCAACTTAAAGTATGGAGTGGC 1327  
Db 601 CTGACAGTCGCTCCTGTGTCAGCGGACGAGTGCAGACACACTGAAGGTGTGGGATGTA 660  
Qy 1328 AGACGAAGAACTGGCACAGGAGCTGCTGGACATGCGGATGAGGTGTTTCGAGTGGACT 1387  
Db 661 AGGCCAGAAAGCTGGCCATGGACCTGCGCGGCCGCGGATGAGGTATATGCTGTTGACT 720  
Qy 1388 GGGCGCCCGATGGCTCTAGAGTTCCTCTGGTGGCAAGGACAAAGTTTATAAGCTATGG 1446  
Db 721 GGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGACAAATGCCTCCGGATATGG 779

RESULT 14

AAH30571  
ID AAH30571 standard; cDNA; 399 BP.  
AC AAH30571;  
XX  
DT 27-JUL-2001 (first entry)  
XX  
DE Human colon cancer cell line Kml2L4-A cDNA library derived sequence #505.  
XX  
KW Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;  
KW detection; colon cancer cell line Kml2L4-A; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200018916-A2.  
PD 06-APR-2000.  
PF 23-SEP-1999; 99WO-US22226.  
XX  
PR 28-SEP-1998; 98US-0102161.  
PR 28-SEP-1998; 98US-0102180.  
PR 29-SEP-1998; 98US-0102380.  
PR 08-OCT-1998; 98US-0103815.  
PR 27-OCT-1998; 98US-0105877.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX  
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX  
DR WPI; 2000-293155/25.  
XX  
PT Polynucleotide library comprising 1079 defined sequences, useful in  
PT the form of an array to detect cancer or susceptibility to cancer -  
XX  
PS Claim 1; Page 328-329; 502pp; English.  
XX

CC The present invention describes a library of polynucleotides comprising  
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described  
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to

CC one of the 1079 sequences; (2) a recombinant host cell containing (I);  
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that  
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method  
CC of detecting differentially expressed genes correlated with a cancerous  
CC state of a mammalian cell comprising detecting a gene product encoded by  
CC 65 of the 1079 sequences given in the specification. The polynucleotides  
CC are used to monitor patients having (or susceptible) to cancer to detect  
CC potentially malignant events at a molecular level before they are  
CC detectable at a gross morphological level. The polynucleotides are also  
CC useful for monitoring the efficacy of various therapies and preventive  
CC interventions. Polynucleotide probes based on the disclosed sequences  
CC are useful for chromosome mapping and detection of transcription levels.  
CC The 1079 polynucleotide sequences were derived from a human colon cancer  
CC cell line Kml2L4-A cDNA library.

SQ Sequence 399 BP; 92 A; 120 C; 112 G; 75 T; 0 other;

Query Match 11.3%; Score 176; DB 21; Length 399;  
Best Local Similarity 66.1%; Pred. No. 3.9e-30;  
Matches 254; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

Qy 266 TCGATATGTGTATCAGCCACAGCGGTTTCAAGTGCGCCAGTGACAAGATGCACGA 325  
Db 16 TAGACATCATCTACCAGCCACAGGCTATCTCAGAGTCCGGGCTGCTGCTGCACCA 75  
Qy 326 GTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTGAATTTACGCCGGATGGTCTC 385  
Db 76 GCTCCTGGAGGTCACAGTGAGGCAGTCATTCTGTGGCCTTCAGCCCTACGGGAAGT 135  
Qy 386 ATCTCGCCAGTGGAAAGTGGCGCACACACAGTCGATTTGGGATCTTAACACAGACACAC 445  
Db 136 ACCTGGCCAGTGGCTCTGGAGACACACACCGCTTCGGGATCTCAGACACAGACACAC 195  
Qy 446 CGCACTTCACTGCACAGGTCATAAGCAGTGGGTTCTGTGCGTATCTCGGGCTCCGGATG 505  
Db 196 CACATTTACATGCAAGGACACAGACACTGGGTCCTTAGTATATCTGTGCTCTCCAGATG 255  
Qy 506 GCAACCGTTGGCCAGCGGTTGCAAGCGGCTCTATATATCATCTGGGACCCCGGAGACGG 565  
Db 256 GCAAGAGAGTGGCCTCAGGCTGCAAGAAATGSCCAGATTCTCCTCTGGGACCCCAAGCAG 315  
Qy 566 GTCAGCAGAAAGGGGCGACCCCTTGAGTGGGCACAAAGAAACACATCAACTGCTCGCTGGG 625  
Db 316 GGAAGCAGGTGGCAGGACCCCTGCGTGGCCACAGCAAGTGGATCACAGGCTGAGCTGGG 375  
Qy 626 AACCGTATCATCGCGATCCGGAGT 649  
Db 376 AGCCCTCCATGCCGAACCTGAGT 399

RESULT 15

AAH05727  
ID AAH05727 standard; cDNA; 547 BP.  
XX  
AC AAH05727;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA clone (5'-primer) SEQ ID NO:2562.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI  
PI Ishii S, Sugiyama T, Wakamatsu A, Naqai K, Otsuki T;  
PI

DR WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

PS Claim 1; SEQ ID 2562; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 547 BP; 108 A; 159 C; 168 G; 109 T; 3 other;

Query Match	10.5%;	Score 164;	DB 22;	Length 547;
Best Local Similarity	62.0%;	Pred. NO. 2.1e-27;		
Matches 274;	Conservative	0;	Mismatches 167;	Indels 1;
				Gaps 1;

QY 54 GATACAGGCGCCTCGTTTACACGGCGGAGGAGCCGGCCGCAATCGACCTGCCGGC 113

Db 80 GCTAGTGCAGTTCAGGATGAGGGCGGCAGCTGCTGGGTCCCCGTTCGACGTGCCCGT 139

OV 114 AGGAATCACTACCCAGCAATTGGGACTGATTTCGAACCGCGCTGCTGAAAAACGAGGAAGC 173

db 140 GGACATCACCCCGACAGGCTGCAGCTCGTGTGCAACGGCGTACTGGCCCGAGGATCC 199

OV 174 CACTCCATATTGTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCTGGAGGACACCGTT 233

200 CCTGCCACCTGGCCCTTTCCTTCCTCAGATCGCTGACACTCCTCATCTCCGACACACCTT 359

**QV**

**23A GGAATTGGGCTCGTCGACCCGAAAAGCTCATAATTTGTCAATCAAGCCCCT** 393

260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

11

QY 474 GTGGTTCTGTGCGTATCCTGG 495  
 ||||| |  
 db 499 CTGGGTCCTTAGGAAATCCNGG 520

Search completed: June 2, 2003, 04:54:48  
Job time : 384 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 04:54:54 ; Search time 101 Seconds  
(without alignments)  
4721.608 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattccccaaaaatgcagga.....aaaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60.2	3.9	2359	1 US-08-188-582-4	Sequence 4, Appli
2	60.2	3.9	2359	1 US-08-646-715-4	Sequence 4, Appli
3	59	3.8	2674	4 US-09-817-180-1	Sequence 1, Appli
4	57.6	3.7	2550	6 5258287-23	Patent No. 5258287
5	57.4	3.7	1221	3 US-08-965-600-2	Sequence 2, Appli
6	57.4	3.7	1221	4 US-09-489-506-2	Sequence 2, Appli
7	56.2	3.6	10136	1 US-08-353-700-2	Sequence 2, Appli
8	56.2	3.6	10136	5 PCT-US95-16216-2	Sequence 1, Appli
9	56	3.6	1872	4 US-09-801-052-1	Sequence 8, Appli
10	55.8	3.6	2085	2 US-08-283-917-8	Sequence 8, Appli
11	55.8	3.6	2085	2 US-08-961-716-8	Sequence 22, Appli
12	55.4	3.6	1736	3 US-09-182-816-22	Sequence 24, Appli
13	55.4	3.6	1736	3 US-09-182-816-24	Sequence 22, Appli
14	55.4	3.6	1736	3 US-09-471-528-22	Sequence 24, Appli
15	55.4	3.6	1736	3 US-09-471-528-24	Sequence 22, Appli
16	55.4	3.6	1736	4 US-09-634-530-22	Sequence 24, Appli
17	55.4	3.6	1736	4 US-09-634-530-24	Sequence 24, Appli
18	55.4	3.6	7218	1 US-08-232-463-14	Sequence 14, Appli
19	55.2	3.5	812	4 US-09-091-097-7	Sequence 7, Appli
20	54.8	3.5	1212	4 US-09-182-145-34	Sequence 34, Appli
21	54.8	3.5	1212	4 US-09-182-145-35	Sequence 35, Appli
22	54.6	3.5	2323	4 US-09-149-476-24	Sequence 24, Appli
23	54.4	3.5	1098	3 US-09-248-335-35	Sequence 35, Appli
24	54.2	3.5	578	4 US-09-602-877A-95	Sequence 95, Appli
25	54.2	3.5	1066	1 US-08-157-101A-4	Sequence 4, Appli
26	53.8	3.5	1117	4 US-09-247-373B-33	Sequence 33, Appli
27	53.6	3.4	6671	1 US-08-280-443-1	Sequence 1, Appli

28	53.6	3.4	6671	1 US-08-457-459-1	Sequence 1, Appli
29	53.6	3.4	6671	1 US-08-555-678-1	Sequence 1, Appli
30	53.6	3.4	6671	5 PCT-US95-02275-1	Sequence 1, Appli
31	53.4	3.4	2184	4 US-08-955-918C-1	Sequence 1, Appli
32	53.4	3.4	2184	4 US-08-697-766A-1	Sequence 1, Appli
33	53.4	3.4	2447	2 US-09-014-969-14	Sequence 14, Appli
34	53.2	3.4	1411	4 US-08-964-127-5	Sequence 5, Appli
35	53.2	3.4	1411	4 US-09-496-692-5	Sequence 5, Appli
36	53.2	3.4	1582	3 US-08-545-196B-10	Sequence 10, Appli
37	53.2	3.4	1582	3 US-08-545-196B-12	Sequence 12, Appli
38	53.2	3.4	3275	4 US-09-370-838-151	Sequence 151, Appli
39	53	3.4	144	1 US-08-702-344-26	Sequence 26, Appli
40	53	3.4	1378	4 US-09-149-476-208	Sequence 208, Appli
41	52.8	3.4	3124	4 US-09-734-030-1	Sequence 1, Appli
42	52.6	3.4	1474	4 US-08-821-994-64	Sequence 64, Appli
43	52.6	3.4	1878	4 US-09-465-558-39	Sequence 39, Appli
44	52.6	3.4	2202	4 US-09-465-558-59	Sequence 59, Appli
45	52.2	3.4	240	1 US-08-628-417-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-08-188-582-4  
; Sequence 4, Application US/08188582  
; Patent No. 5534410  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlact, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; NUCLEIC ACIDS ENCODING TAFS AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/188,582  
; FILING DATE: 28-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2359 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 49..2160





```
; LENGTH: 2674
; TYPE: DNA
; ORGANISM: Human
US-09-817-180-1

Query Match      3.8%; Score 59; DB 4; Length 2674;
Best Local Similarity 55.0%; Pred. No. 1.4e-06;
Matches 116; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1345 CAGGAGCTGCTGGACATGCGGATGAGGTGTTCCGGAGTGGACTGGCGCCCGATGGCTCT 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2439 CAGCATCCACACTGCCGGCAGGATGCAGGCCGCTGTCTCTGTGTCTCTGTCTCTGTCT 2498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1405 AGAGTTGCCTCTGTTGGCAAGGACAAAGTTATATAAGCTATGGCTTATTAAACAAATCATT 1464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2499 AGGGCTTCCTCTCCGGGCAGAAACAATAAAACCACTTGTGCCACTGTAAGAAAAA 2558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1465 AACTGTACACGGTAAGAAATACCTTAGGAATAAAGTAAACGTCCTGTAGTAAAAA 1524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2559 AAAAAA 2618

QY 1525 AAAAAA 1555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2619 AAAAAA 2649
```

```
RESULT 4
5258287-23
; Patent No. 5258287
; APPLICANT: BAXTER, ROBERT C.; WOOD, WILLIAM I.
; TITLE OF INVENTION: DNA ENCODING AND METHODS OF PRODUCTION
; OF INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN BP53
; NUMBER OF SEQUENCES: 58
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/171,623
; FILING DATE: 22-MAR-1988
; SEQ ID NO: 23:
; LENGTH: 2550
5258287-23
```

```
Query Match      3.7%; Score 57.6; DB 6; Length 2550;
Best Local Similarity 65.6%; Pred. No. 3.4e-06;
Matches 84; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1426 GACAAAGTTATAAAGCTATGGGCTTATTAAACAAATCATTAACCTGTACACGTAAGAAA 1485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2423 GAGAAAGTTGTCTGCAATGTAATTATATAAATAGTAAATAGTTTACCATTAAAAA 2482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1486 TACTTAGGAATAAAGTAAACGTCCTGAGTAAAAA 1545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2483 AAAAAA 2542

QY 1546 AAAAAA 1553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2543 AAAAAA 2550
```

```
RESULT 5
US-08-965-600-2
; Sequence 2, Application US/08965600
; Patent No. 6077688
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW TRANSDUCIN BETA-1 SUBUNIT
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
```

```
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/965,600
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0416 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1221 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KIDNNOT02
; CLONE: 194046
US-08-965-600-2
```

```
Query Match      3.7%; Score 57.4; DB 3; Length 1221;
Best Local Similarity 48.6%; Pred. No. 2.6e-06;
Matches 157; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 1130 TGAATATTCCCGGATGTAAAGCTAATTCGCTCTGCTTCATTTGACAAGTCAGTGGGTC 1189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 666 TGACCTTTCCCGGACTCCAGCTCCTTGCTCACTGCTTCAGATGATGGCTACATCAAGA 725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1190 TGTGGCAGCCAGCGGATGGTCAGTACATGGCCACCTCCGGGGTTCATGTCAGGCTGTTT 1249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 726 TCTATGATGTACAACATGCCAATTTGGCTGGCAGCTGAGCGGCCATGCTCCTGGGTGC 785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1250 ACACGGTTGCCTGGTCCGCGGACTCCCGCTTGATTGTTCCGGCAGCAAGACTCAATC 1309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 786 TGAACGTTGCTCTGCTGTGATGACACTCACTTTGTTTCCAGTTCGCTGACAAAAAGTG 845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1310 TAAAGTATGGAGTGTGCAGACGAGAAACTGGCAGAGAGCTGCCTGGACATGCGGATG 1369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 846 TAAAGTTGGGATGTTGGAACGAGGACTTGTGTTACACACTTCTTTGATCACCAGGATC 905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1370 AGGTGTTCCGAGTGGACTGGGCGCCCGATGGCTCTAGAGTTGCCCTGTTGGCAAGACA 1429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 906 AGGTCTGGGAGTAAATACAAATGGAATGGTTCAAAATTTGTCTGTTGGAGATGACC 965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1430 AAGTTATAAGCTATGGGCTTAT 1452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 966 AGGAAATTCACATCTATGATTGT 988
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 6
US-09-489-506-2
; Sequence 2, Application US/09489506
; Patent No. 6465619
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW TRANSDUCIN BETA-1 SUBUNIT
; NUMBER OF SEQUENCES: 3
```

/  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 / STREET: 3174 Porter Drive  
 / CITY: Palo Alto  
 / STATE: CA  
 / COUNTRY: USA  
 / ZIP: 94304  
 /  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Diskette  
 / COMPUTER: IBM Compatible  
 / OPERATING SYSTEM: DOS  
 / SOFTWARE: FastSEQ for Windows Version 2.0  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/09/489,506  
 / FILING DATE:  
 / CLASSIFICATION:  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/965,600  
 / FILING DATE:  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Billings, Lucy J,  
 / REGISTRATION NUMBER: 36,749  
 / REFERENCE/DOCKET NUMBER: PF-0416 US  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 650-855-0555  
 / TELEFAX: 650-845-4166  
 / TELEX:  
 / INFORMATION FOR SEQ ID NO: 2:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 1221 base pairs  
 / TYPE: nucleic acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / IMMEDIATE SOURCE:  
 / LIBRARY: KIDNNOT02  
 / CLONE: 194046  
 /  
 / US-09-489-506-2

Query Match	3.7%;	Score 57.4;	DB 4;	Length 1221;
Best Local Similarity	48.6%;	Pred. No. 2.6e-06;		
Matches 157; Conservative	0;	Mismatches 166;	Indels	0

QY	1130	TGAAATATTCCCGCGGATGTAAAGCTAATTGGCTCTGCCTTCATTTGACAAAGTCAAGTGCGTC	1189
Dd	666	TGACCTTTTCCCCGGACTCCCAGCTCCTTGTCACTGCTTCAGATGATGGCTACATCAAGA	725
QY	1190	TGTGGCGAGCCAGCATGSTCAGTACATGCCACCCTTCGGGGGTCAATGTCAGGCGTGT	1249
Dd	726	TCTATGATGTACAACATGCCAATTGGCTGGCACGCTGAGCGGCCATGCCTCCTGGGTGC	785
QY	1250	ACACGGTTGCCTGGTCCGGGACTCCCCTTGATTGTTTCCGGCAGCAAAGACTCAACTC	1309
Dd	786	TGAACGTTGCATTCTGTCTGTATGACACTCACTTTGTTHCCAGTTCGTCTGACAAAAGTG	845
QY	1310	TAAAGTATGGAGTGTGCAGACGAAGAAACTGGCACAGSAGCTGCCCTGGACATGCGGATG	1369
Dd	846	TAAAGTTTGGGATGTTGGAACGAGGACTTGTGTTACACACCTTCTTTGATCACCAGGATC	905
QY	1370	AGGTGTTCCGGAGTGGACTGGCGGCCCGATGGCTCTAGAGTTGCCCTCGGTGGCAAGGACA	1429
Dd	906	AGGTCTGGGGAGTAATAACAATGGAATGGTTCAAAAATTGTCTGTTGGAGATGACC	965
QY	1430	AAGTTATAAGCTATGGCTTAT	1452
Dd	966	AGGAAATTACATCTATGATTGT	988

RESULT 7  
US-08-353-700-2  
; Sequence 2, Application US/08353700  
; Patent No. 5599919  
; GENERAL INFORMATION:  
; APPLICANT: YEN, TIMOTHY J.

```

; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
;
US-08-353-700-2

Query Match          3.6%; Score 56.2; DB 1; Length 10136;
Best Local Similarity 62.4%; Pred. No. 1.7e-05;
Matches 88; Conservative 0; Mismatches 53; Indels 0;

QY      1415 CTGCTGGCAAGGACAAAGTTATAAAGCTATGGGCTTATTAAACAATCATTAAC
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db      9990 CTGTTAGCATTTGCCATTCTCTACTGCAATGTAATAAGTATGTAATATGTATA
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY      1475 CGGTAAGAAAATACATTAGGAATAAAGTAAACGTCCTGAGTAAAAA
      | | | | ||| | ||| | ||| | ||| | ||| | ||| |
Db      10050 TTTTGGTAATATGTTACAATTAAATGCAAGCACTATATAAAAAA
      | | | | ||| | ||| | ||| | ||| | ||| | ||| |
QY      1535 AAAAAAAAAAAAAAAAAAAAAA 1555
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      10110 AAAAAAAAAAAAAAAAAAAAAA 10130

RESULT 8
PCT-US95-16216-2
; Sequence 2, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307

```



Db 1391 AATGCATCAGAACCAATGCATGGCCATGACCAACAATGTTCTTCAGTAGCCATCATGCCCA 1450  
QY 1145 ATGTAAAGCTAATTCGCTGCTCTTCAATTTGACAAGTCAGTGGTCTGTGGCGAGCCAGG 1204  
Db 1451 ATGGAGATCATATAGTGTCTGCTCAAGGGATAAAACTATAAAATGTGGGAAGTGCAA 1510  
QY 1205 ATGGTCAGTACATGGCCACTTCCGGGTCATGTGCAGGCTGTTTACACGGTTGCCTGGT 1264  
Db 1511 CTGGCTACTGTGTGAAGACATTCACAGACACAGAGAATGGTACGTATGGTGGGCCAA 1570  
QY 1265 CCGCGGACTCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAAAGTATGGAGTG 1324  
Db 1571 ATCAAGACGGCACTCTGATAGCCAGCTGTTCCAATGACCAAGACTGTGCGTGTATGGTGC 1630  
QY 1325 TGCAGACGAAGAACTGGCACAGGAGTGCCTGGACATGCGGATGAGGTGTTTCGGAGTGG 1384  
Db 1631 TAGCAACAAGGAATGCAAGGCTGAGCTCGAGAACATGAGCATGTGGTAGAATGCATTT 1690  
QY 1385 ACTGGGCGCCCGATGGCTC 1403  
Db 1691 CCTGGGCTCCTGAAAGCTC 1709

RESULT 11  
US-08-961-716-8  
; Sequence 8, Application US/08961716  
; Patent No. 5880272  
; GENERAL INFORMATION:  
; APPLICANT: ADACHI, HIDEKI  
; APPLICANT: TSUJIMOTO, MASAFUMI  
; APPLICANT: INOUE, KEIZO  
; APPLICANT: ARAI, HIROYUKI  
; TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME  
; TITLE OF INVENTION: AND GENE THEREOF  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER &  
; ADDRESSEE: NEUSTADT,P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,716  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/283,917  
; FILING DATE: 03-AUG-1994  
; APPLICATION NUMBER: JP 209943/1993  
; FILING DATE: 03-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5880272man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 2292-030-0  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; TELEX: 24855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2085 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: cdna

; ORIGINAL SOURCE:  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 844..2073  
US-08-961-716-8  
Query Match 3.6%; Score 55.8; DB 2; Length 2085;  
Best Local Similarity 46.7%; Pred. No. 9.6e-06;  
Matches 177; Conservative 0; Mismatches 202; Indels 0; Gaps 0;  
QY 1025 AGTCGCTGGTTTCTCTGTTTCGGATGACAACACCCCTCTATCTGTGGCGGAACAACAGACA 1084  
Db 1331 AGCTTCTGGCTTCATGTTCTGCAGATATGACCATTAAGCTATGGGATTTTCAGGGCTTTG 1390  
QY 1085 AGTGGCTTGAGCGCATGACAGGGCACAGAGCGTGGTCAACGATGTGAATATTCGCCGG 1144  
Db 1391 AATGCATCAGAACCATGCATGGCCATGACCAATGTTTCTCAGTAGCCATCATGCCCA 1450  
QY 1145 ATGTAAAGCTAATTCGCTCTGCTTCATTTTGACAAGTCAGTGGCTCTGTGGCGAGCCAGG 1204  
Db 1451 ATGGAGATCATATAGTGTCTGCTCAAGGGATAAAACTATAAAATGTGGGAAGTGCAA 1510  
QY 1205 ATGGTCAGTACATGGCCACTTCCGGGTCATGTGCAGGCTGTTTACACGGTTGCCTGGT 1264  
Db 1511 CTGGCTACTGTGTGAAGACATTCACAGACACAGAGAATGGTACGTATGGTGGGCCAA 1570  
QY 1265 CCGCGGACTCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAAAGTATGGAGTG 1324  
Db 1571 ATCAAGACGGCACTCTGATAGCCAGCTGTTCCAATGACCAAGACTGTGCGTGTATGGTGC 1630  
QY 1325 TGCAGACGAAGAACTGGCACAGGAGTGCCTGGACATGCGGATGAGGTGTTTCGGAGTGG 1384  
Db 1631 TAGCAACAAGGAATGCAAGGCTGAGCTTCGAGAACATGAGCATGTGGTAGAATGCATTT 1690  
QY 1385 ACTGGGCGCCCGATGGCTC 1403  
Db 1691 CCTGGGCTCCTGAAAGCTC 1709  
RESULT 12  
US-09-182-816-22  
; Sequence 22, Application US/09182816  
; Patent No. 6143542  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: LO, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1  
; CURRENT APPLICATION NUMBER: US/09/182,816  
; CURRENT FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (159)..(1553)  
US-09-182-816-22  
Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGATTATAAAGCTATGGGTTATTAAACAATCATTAATTAATTTGTACACGGTAAGAAAT 1486



Db 1588 ATAATGTTAAAAATAAATGTAATTACTGCTGAAATAAACGATATGGATTTTATTTCAAACT 1647  
QY 1487 ACTTAGGAATAAAGTAAACGCTCTGAGTAAACAAAAAAGAAAAAAGAAAAAAGAAAAA 1546  
Db 1648 TGTCAAATATAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 1707  
QY 1547 AAAAAAAA 1555  
Db 1708 AAAAAAAA 1716

RESULT 13  
US-09-182-816-24/c  
; Sequence 24, Application US/09182816  
; Patent No. 6143542  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1  
; CURRENT APPLICATION NUMBER: US/09/182,816  
; CURRENT FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-182-816-24

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAAGCTATGGCTTATTAAACAATCATTAACCTGTACACGGTAAGAAAT 1486  
Db 149 ATAATGTTAAAAATAAATGTAATTACTGCTGAAATAAACGATATGGATTTTATTTCAAACT 90  
QY 1487 ACTTAGGAATAAAGTAAACGCTCTGAGTAAACAAAAAAGAAAAAAGAAAAAAGAAAAA 1546  
Db 89 TGTCAAATATAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 30  
QY 1547 AAAAAAAA 1555  
Db 29 AAAAAAAA 21

RESULT 14  
US-09-471-528-22  
; Sequence 22, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1736

; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (159)..(1553)  
US-09-471-528-22  
Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAAGCTATGGCTTATTAAACAATCATTAACCTGTACACGGTAAGAAAT 1486  
Db 1588 ATAATGTTAAAAATAAATGTAATTACTGCTGAAATAAACGATATGGATTTTATTTCAAACT 1647  
QY 1487 ACTTAGGAATAAAGTAAACGCTCTGAGTAAACAAAAAAGAAAAAAGAAAAAAGAAAAA 1546  
Db 1648 TGTCAAATATAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 1707  
QY 1547 AAAAAAAA 1555  
Db 1708 AAAAAAAA 1716

RESULT 15  
US-09-471-528-24/c  
; Sequence 24, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-471-528-24

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAAGCTATGGCTTATTAAACAATCATTAACCTGTACACGGTAAGAAAT 1486  
Db 149 ATAATGTTAAAAATAAATGTAATTACTGCTGAAATAAACGATATGGATTTTATTTCAAACT 90  
QY 1487 ACTTAGGAATAAAGTAAACGCTCTGAGTAAACAAAAAAGAAAAAAGAAAAAAGAAAAA 1546  
Db 89 TGTCAAATATAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 30  
QY 1547 AAAAAAAA 1555  
Db 29 AAAAAAAA 21

Search completed: June 2, 2003, 06:54:25  
Job time : 104 secs





Db 190 CTTGCCACTGGCTTTCTTTGTCCACGATGCTGAGATCGTCTCCTCACTGGGGAAGACGTT 249  
Qy 234 GGAATTGGCTCAGTGGACACCGAAGACGTGATCGATATTGTGTATCAGCCACAGGCGGT 293  
Db 250 GGAGTCCCAGGAGTGAGACAGAGAAGGTCTTAGACATCATCTACCAGCCACAGGCTAT 309  
Qy 294 TTTCAAAGTGGCCCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGT 353  
Db 310 CTTCAGAGTCCGGGCTGTGACTCGCTGCACCAAGCTCCTTGGAGGTCACAGTGAGGCAGT 369  
Qy 354 GGTTCGCTGAATTTACGCCCGGATGCTGCTCATCTCGCCAGTGGAAAGTGGGACACCCAC 413  
Db 370 CATTTCTGTGGCCTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACCCAC 429  
Qy 414 AGTGCAGATTGTGGGATCTTAACACAGACACCGCACTTCACCTGCACAGGTCATGAAGCA 473  
Db 430 CGTGCCTTCTGGGATCTCAGCACAGACACACCACTATTTCACATGCAAGGGACACAGACA 489  
Qy 474 GTGGGTTCTGTGCGTATCCTGGGCTCCGATGGCAACGGTTGGCCAGCGGTGCAAGC 533  
Db 490 CTGGGTCCTTAGTATATCCTGGTCTCCAGATGGCAAGACTGGCCTCAGGCTGCAAGAA 549  
Qy 534 GGGCTCTATATCATCTGGACCCCGAGACGGGTACGACAGAAGGGCGACCCCTGAGTGG 593  
Db 550 TGSCCAGATTCTCCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGSCAGGACCCCTCGCTGG 609  
Qy 594 GCACAAAGAACACATCAACTGCCTCGCTGGGAACCGTATCATCGGATCCGGAGTGCAG 653  
Db 610 CCACAGCAAGTGGATCACAGGCCTGAGCTGGGAGCCCTCCATGCGAAACCCCTGAGTGCCG 669  
Qy 654 GAAACTTGCTCCCGCAGTGGAGACGGGACTGCCGATTTGGACGTAATAATGGGCCA 713  
Db 670 CTATGTGGCCAGCAGCTCCAAGGATGGCAGTGTGCGATCTGGGACACAATGCAAGGCCG 729  
Qy 714 GTGCCTTATGACATTGCCCGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGC 773  
Db 730 CTGTGAGCGCATCTCACCGGGCACACCCAGTCCGTGCTCACCTGTCTCCGGTGGGAGGGGA 789  
Qy 774 GGGCCTTATTATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG 833  
Db 790 CGGCCTTCTCTACTCTGCCTCCAGGACCGGCACCACTCAAAAGTCTGGAGAGCTCATGACGG 849  
Qy 834 AATCTTGTCCCGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAAACATTTGCGTGAACAC 893  
Db 850 TGTGCTGTCCCGGACTCTGCAAGGCCACCGGCCACTGGGTGAACACCATGCGCCCTCAGCAC 909  
Qy 894 CGATTACGTCCTGCGCACATGGTCCATTCATCCGGTGAAGGATCGCTCCAAGACC---A 950  
Db 910 TGACTATGCCCTGCGCACATGGGGCCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCCAAGA 969  
Qy 951 CTTCAAGTTTGACCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGSCCGTGTG 1010  
Db 970 CCTCAAGGATCCTTGCAGGAGTTGAAGGAGAGGGCTCTGAGCCGATACAACCTCGTGCG 1029  
Qy 1011 CCCTGACGAGTGGAGTCCGCTGTTTCCCTGTTCCGATGACAACACCCCTCTATCTGTGG-- 1068  
Db 1030 GGGCCAGGTTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACTTATTCCTGTGGTC 1089  
Qy 1069 -CGAACAACAGAAAGTGCCTTGAAGCGCATGACAGGGCACCAAGAACGTTGGTCAACGA 1127  
Db 1090 CCCAGCAGAGGACAAAAGCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA 1149  
Qy 1128 TGTGAATATTGCGCCGATGTAAGCTAATTTGCGTGTGCTTCAATTTGACAAGTCAGTGCG 1187  
Db 1150 GGTGNTCTTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCCTTTGACAAGTCCATCAA 1209  
Qy 1188 TCTGTGGGAGCCAGCGATGGTCACTACATGGCCACCTTCCGGGTCATGTGCGAGGCTGT 1247  
Db 1210 GCTGTGGGATGCGAGGACGGGCAAGTACCTGGCTTCCCTACGCGGCCACGTTGCGCTGCCGT 1269  
Qy 1248 TTACACGGTTCGCTGGTCCCGGACTCCCGCTTGATGTTTCCGGCAGCAAAAGACTCAAC 1307

Db 1270 GTACCAGATTGCGTGGTTCAGCTGACAGTCCGCTCCTGGTCAAGCGCAGCAGTGCACAGCAC 1329  
Qy 1308 TCTAAAAGTATGGAGTGTGCAGACGAAAGAACTGGCACAGGAGCTGCCTGGACATGCCGA 1367  
Db 1330 ACTGAAGGTGTGGGATGTGAAGGCCCAAGAACTGGCCATGACCTGCCCGCCACCGCGGA 1389  
Qy 1368 TGAGGTGTTCCGAGTGGACTGGCGCCCGATGGCTCTAGAGTTGCTCTGTGGTGCACAAGGA 1427  
Db 1390 TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGATGGCAAGTGGTGGGAAGGA 1449  
Qy 1428 CAAAGTTATAAAGCTATGG 1446  
Db 1450 CAAATGCCCTCCGGATATGG 1468

RESULT 2  
US-10-132-744A-1  
; Sequence 1, Application US/10132744A  
; Publication No. US20030027261A1  
; GENERAL INFORMATION:  
; APPLICANT: Utku, Nalan  
; TITLE OF INVENTION: No. US20030027261alel genes Tzap7/A, Tzap7/B and Tzap7 involv  
; TITLE OF INVENTION: activation and uses thereof  
; FILE REFERENCE: Utku-4 CON  
; CURRENT APPLICATION NUMBER: US/10/132,744A  
; CURRENT FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: PCT/EP00/10670  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 60/185,016  
; PRIOR FILING DATE: 2000-02-25  
; PRIOR APPLICATION NUMBER: 60/162,675  
; PRIOR FILING DATE: 1999-11-01  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1928  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (92)..(1543)  
US-10-132-744A-1

Query Match 30.0%; Score 466.2; DB 9; Length 1928;  
Best Local Similarity 59.3%; Pred. No. 5.1e-115;  
Matches 830; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

Qy 54 GATACAGGGCGGCTCGTTTACACGGGCGAGGAAGCGCGCCGCAATCGACCTGCCGCG 113  
Db 139 GCTAGTGCAGTCCAGGATGAGGGCGGSCAGCTGCTGGGTCCCGCTTCGACGTCGCCGT 198  
Qy 114 AGGAATCACATACCCAGCAATTGGGACTGATTTGCAACGGCTGCTGAAACACGAGGAAGC 173  
Db 199 GGACATCACCCCGGACAGGCTGCAGCTGTTTGCAACGGCTACTTGGCCCGCAGGAGATCC 258  
Qy 174 CACTCCATATTGTTTTCGTGGGCGAGGATGAGATCAAGAAGAGCCTGGAGACACGTT 233  
Db 259 CCTGTCACTGGCTTCTTGTCCACGATGCTGAGATCGTTTCCACTCGGGAAGACGTT 318  
Qy 234 GGAATTGGCGTCAGTGGACACCGAAACGTCGATCGATATTGTGTATCAGCCACAGGCGGT 293  
Db 319 GGAGTCCCAGCAGTGGACAGAGAGGTCCTAGACATCATTTTACCCACACAAGCTGT 378  
Qy 294 TTTCAAAGTGGCCCCAGTGACAAGATGCACGAGTTCATGCCGGACACGCCGAGGCTGT 353  
Db 379 GTTCAAAGTTCGTGCTGTAAACAGATGTACCAGCTCATTGGAGGACACACCCGAGGCTGT 438  
Qy 354 GGTTCGCTGATTTTCAGCCCCGGATGGTGTCTCATCTGCCAGTGGAAAGTGGCGACACCAC 413  
Db 439 TATTTCACTAGCCTTCAGCCCAACTGGAAAGTATTTGGCAAGTGGTTCTTGGGACACTAC 498  
Qy 414 AGTGCAGTTGCGGATCTTAACACAGACACCGCAGTTCACCTGCACAGGTCATGAAGCA 473



Db 499 AGTCGCGCTTTGGGATCTCAGCACAGAACTCCACATTTTACATCTAAAGGGGCATACACA 558

QY 474 GTGGGTTCTGTGCGTATCCTGGCTCGGATGGCAACAGGTTGGCCAGCGGTTGCCAAAGC 533

Db 559 CTGGGTTCTCAGTATTGCTTGGTCTCCAGATGGCAAAACTTGCCTCAGGATGTAAGAAA 618

QY 534 GGGCTCTATAATCATCTGGGACCCGGAGACGGGTGACAGAGGGGGACCCCTTGAGTGG 593

Db 619 TAGTCAGATCTTCATTTGGGACCCCAAGCACAGGGAAGCAGATTGGCAAACTTAACAGG 678

QY 594 GCACAAGAAACACATCACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAG 653

Db 679 GCACTCAAAGTGGATTACATGGCTGTGTGGGAACCTCTCCACCTGAACCCACAGAGCCG 738

QY 654 GAAACTTGCTTCGCCAGTGGAGACGGGACTGCCGGATTGGGACGTAATAATTGGGCCA 713

Db 739 ATACCTAGCCAGTGCCTCCAGCGCGCGCTCGACCGGATCTGGACACAACACTGCAGGCCG 798

QY 714 GTGCTTATGAACATTCGCGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGC 773

Db 799 CTGTGAGCGCATCCTCACCGGCACACCCAGTCGGTCACTGTCTCCGGTGGGAGGGGA 858

QY 774 GGGCCTTATTATACATCTCCAAAAGATCGCACAGTGAAGATGTGGCAGCAGCTGATGG 833

Db 859 CGGCTTCTCTACTCTGCTCCAGGACCGCCACTGGGTGAACACCATGGCCCTCATGACGG 918

QY 834 AATCTTGTGCGGACGTTCTCTGGCCAAAGCTCACTGGGTAACAACTGCGCTGAGCAC 893

Db 919 TGTGCTGTGCGGACTCTGCAAGGGCCACGGCCACTGGGTGAACACCATGGCCCTCAGCAC 978

QY 894 CGATTACGTCCTGCGCACTGTFPCATTCCATFCCGGTGAAGGATCGCTCCAAGAGCC---A 950

Db 979 TGACTATGCCCTGCGCACTGGGGCCTTTGAACCTGCTGAGGCCTCAGTTAATCCCCAAGA 1038

QY 951 CCTCAGTTTGAGCACTGAGGAATTGCAGGAATCTGCCCTTGAGCGCTACAGGCCGTGTG 1010

Db 1039 CTTCCAAGGATCCTTGCAAGGAGTTGAAGGAGGGCTCTGAGCCGATACAACTCGTGGC 1098

QY 1011 CCTGACGAGGTGGAGTGGTGGTTTCTGTTCCGGATGACAAACACCCCTCTATCTGTGG-- 1068

Db 1099 GGGCAGGGTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACCTTATTCCTGTGTGTC 1158

QY 1069 -CGGAACAACCAAGTGGCTGAGCGCATGACAGGGCCACAGAACCGGTCAACGA 1127

Db 1159 CCCAGCAGAGGACAAAGCCCTCTCACTCGGATGACAGGACCAAGCTCTCATCAACCA 1218

QY 1128 TGTGAAATATTCGCCGGATGTAAGCTAATTGGCTCTGCTTCATTTGACAACTCAGTGC 1187

Db 1219 GGTGCTCTTCTCTCTGACTCCCGCATCGTGGCTAGTGGCTCCTTTGACAACTCCATCAA 1278

QY 1188 TCTGTGGCGAGCCAGCGATGGTCACTACATGGCCACCTTCGGGGTTCATGTGAGGCTGT 1247

Db 1279 GCTGTGGGATGGCAGACGGGCAAGTACCTGGCTTCCCTACCGGGGCCACGTGGCTGCCGT 1338

QY 1248 TTACACGGTTGCTGGTCCGCGGACTCCCGCTTGATTTTCCGGCAGCAAAAGACTCAAC 1307

Db 1339 GTACCAGATTGCGTGGTCAGCTGACAGTCGGCTCCTGGTCAAGCGGACGAGTGACAGCAC 1398

QY 1308 TCTAAAAGTATGGAGTGTGCAGACCAAGAAACTGGCACAGGAGTGCCTGGACATGCGGA 1367

Db 1399 ACTGAAGGTGTGGGATGTGAAGGCCAAGAACTGGCCATGGACCTGCCCCGCCAGCGGA 1458

QY 1368 TGAGGTGTTCGGAGTGGACTGGGCGCCGATGGCTCTAGAGTTGCTTGGTGGCAAGGA 1427

Db 1459 TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA 1518

QY 1428 CAAAGTTATAAAGCTATGG 1446

Db 1519 CAAATGCCCTCCGGATATGG 1537

RESULT 3

US-10-132-744A-3

; Sequence 3, Application US/10132744A

; Publication No. US20030027261A1

; GENERAL INFORMATION:

; APPLICANT: Utiku, Nalan

; TITLE OF INVENTION: No. US20030027261A1el genes Tzap7/A, Tzap7/B and Tzap7 involve

; TITLE OF INVENTION: activation and uses thereof

; FILE REFERENCE: Utiku-4 CON

; CURRENT APPLICATION NUMBER: US/10/132,744A

; CURRENT FILING DATE: 2002-04-24

; PRIOR APPLICATION NUMBER: PCT/EP00/10670

; PRIOR FILING DATE: 2000-10-30

; PRIOR APPLICATION NUMBER: 60/185,016

; PRIOR FILING DATE: 2000-02-25

; PRIOR APPLICATION NUMBER: 60/162,675

; PRIOR FILING DATE: 1999-11-01

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3

; LENGTH: 1170

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (3)..(785)

US-10-132-744A-3

Query Match 16.0%; Score 249.4; DB 9; Length 1170;

Best Local Similarity 59.3%; Pred. No. 7.6e-57;

Matches 462; Conservative 0; Mismatches 311; Indels 6; Gaps 2;

QY 674 GAGACGGGGACTGCCGGATTTTGGGACGTAAATTTGGGCCAGTGCCTTATGAACATTCGCG 733

Db 1 GCGGCGCGCTCGACCGGATCTGGGACACAACCTGCAGGCCGCTGTGAGCGCATCTCACC 60

QY 734 GACACACAAATGCTGTGACAGCAGTGAAGATGGGTGGAGCGGCGCTTATATACATCCT 793

Db 61 GGCACACCCAGTCGGTCACTGTCTCCGGTGGGAGGGGACGGGCTTCTCTACTCTGCCT 120

QY 794 CCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTGTGTGCCGACGTTCT 853

Db 121 CCCAGACCGCCACCATCAAAAGTCTGGAGAGCTCATGACGGTGTGCTGTGCCGACTCTGC 180

QY 854 CTGGCCAGCTCACTGGGTAAACAACTTGGCTGAGCACCAGTACCTTACGTCCTGCCGACTG 913

Db 181 AAGGCCACGGCCACTGGGTGAACACCATGGCCCTCAGCAGCTGACTATGCCCTGCCGACTG 240

QY 914 GTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCC---ACCTCAGTTTGAGCAGCTGAGG 970

Db 241 GGGCCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCCAAGACCTCCAGGATCCTTTGCAGG 300

QY 971 AATTGAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTGCCCTTGACGAGGTGGAGTGC 1030

Db 301 AGTTGAAGAGAGGGCTCTGAGCCGATACAACTCTCGTGGGGGCCAGSGTCCAGAGAGGC 360

QY 1031 TGGTTTCCCTGTTCCGATGACAAACACCCCTCTATCTGTGG--CGGAACAACCAAGCAAGT 1087

Db 361 TGGTGTCTGGCTCCGACGACTTCACCTTATTCCTGTGTGTCCTCCAGCAGAGGACAAAAAGC 420

QY 1088 GCGTTGAGCGCATGACAGGGCACCAAGACCTGGTCAACGATGTGAAATATTCGCCGATG 1147

Db 421 CTCTCAGTCGGATGACAGGACCAAGCTCTCATCAACCAAGTCCATCAAGCTGTCTCTCTGACT 480

QY 1148 TAAAGCTAATGCGTCTGCTTCATTTGACAAGTCAAGTGCCTGTGTGGGAGCCAGCGATG 1207

Db 481 CCCGCATCGTGGCTAGTGCCTCTCTTTGACAAGTCCATCAAGCTGTGGGATGGCAGGACGG 540

QY 1208 GTCAGTACATGGCCACCTTCGGGGGTCAATGTCAGGCTGTTTACACGGTGCCTTGGTCCG 1267

Db 541 GCAAGTACCTGGCTTCCCTACGCGGCCACCTGCGGTGCGGTGTACCAGATTGCGTGGTCA 600

QY 1268 CGGACTCCCGCTTGATTGTTTCCGGCAGCAAAAGACTCAACTCTAAAAGTATGGAGTGTGC 1327

Db 601 CTGACAGTCGGCTCCTGGTCAAGGGCAGCAGTGAAGGTGTGGGATGTGA 660





Db 63 GAGCTCATGACGGTGTGCTGTGCGGACTCTGCAAGGCCACGGCCACTGGGTGAACACCA 122  
QY 881 TTGGCGCTGAGCACCAGATTAGCTCTGGCAGCTGGTCCATTCCATCCGGTGAAGGATCGCT 940  
Db 123 TGGCCCTCAGCACTGACTATGCGCTGGCAGCTGGGGCCCTTTGAACCTGCTGAGGCCTCAG 182  
QY 941 CCAAGAGCC---ACCTCAGTTTGAGCAGCTAGGAATTCAGGAATCTGCCTTGAAGCGCT 997  
Db 183 TTAATCCCAAGACCTCCAAGATCCTTGACAGAGTTGAAGAGAGGGCTCTGAGCCGAT 242  
QY 998 ACCAGGCCGTGTGCCCTGACAGGTGGAGTCGCTGGTTTCCTGTTCCGATGACAAACACC 1057  
Db 243 ACRACCTCGTGGCGGGCCAGGTCAGAGAGGCTGGTGTCTGGCTCCGACACTTCACCT 302  
QY 1058 TCTATCTGTGG---CGGAACACACAGAACAAAGTCGTTGACGCGCATGACAGGGCACCA 1114  
Db 303 TATTCCTGTGTTCCCGCAGCAGGACAAAAGCCCTCTCACTCGGATGACAGGACACCAAG 362  
QY 1115 ACGTGGTCAACGATGTGAATATTCGCCGGATGTAAGCTAATTCGCTGCTTCATTG 1174  
Db 363 CTCTCATCAACCAAGGTGCTCTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCCTTTG 422  
QY 1175 ARAAGTCAGTGGCTGTGG 1194  
Db 423 ARAAGTCCATCAAGCTGTGG 442

RESULT 8  
US-09-764-853-29  
; Sequence 29, Application US/09764853  
; Patent No. US20020090672A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJ206  
; CURRENT APPLICATION NUMBER: US/09/764,853  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 939  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 3025  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (3011)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (3014)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (3015)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-853-29

Query Match 5.1%; Score 79.4; DB 10; Length 3025;  
Best Local Similarity 49.2%; Pred. No. 6.5e-11;  
Matches 209; Conservative 0; Mismatches 216; Indels 0; Gaps 0;  
QY 1025 AGTCGCTGTTTCTCTGTCGGATGACAAACCCCTCTATCTGTGGCGGAACACCAACA 1084  
Db 222 AGCAACTGCTACTGCTTCTTGGGATACCTTTCTCATGCTATGGAATTTCAAGCCACATG 281  
QY 1085 AGTCGCTGAGCGCATGACAGGGCACCAGAACGTGGTCAACGATGTGAATATTCGCCGG 1144  
Db 282 CTAGAGCTTACAGATATGTGGGTCAAGGATGTTTAACCAAGCGTGCAGTTTCTCCAC 341  
QY 1145 ATGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTCCGCTGTGTGGCGGACGCG 1204  
Db 342 ATGGAAGCTATTGGCGTCTGCTCTCAGGACAGACAGACCGTGAGACTCTGGATTCTTGATA 401

QY 1205 ATGGTCAGTACATGGCCACCTTCCGGGGTCATGTCAGGCTGTTTACACGGTTCGCTGGT 1264  
Db 402 AGAGAGGAAAATTTCTAGAATTTAAAGCTCATACAGCTCCAGTTCGAAAGTGTAGACTTTT 461  
QY 1265 CCGCGGACTCCCGCTGATTTGTTCCGGCAGCAAGAGACTCAACTCTAAAAAGTATGGAGTG 1324  
Db 462 CAGCTGATGGCCAGTTCTAGCTACAGCTTCTGAAGACAAATCCATAAAAGTATGGAGCA 521  
QY 1325 TGCAGACGAAGAACTGGCACAGGAGCTGCCCTGGACATCGCGGATGAGGTGTTCCGAGTGG 1384  
Db 522 TGTATCGCCAGCGCTCCTGTATTCTCTGTATCGACATACACACTGGGTACGCTGTGCCA 581  
QY 1385 ACTGGCGCCCGATGGCTCTAGAGTTGCCCTCTGCTGGCGAAGGACAAAGTTATAAAGCTAT 1444  
Db 582 AATTTCAACCCGATGGAAGACTAATTGTGTCTATGTAGTGAGGATAAAACTATTAAATTT 641  
QY 1445 GGGCT 1449  
Db 642 GGGAT 646

RESULT 9  
US-09-794-257-4  
; Sequence 4, Application US/09794257  
; Patent No. US20020009804A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US20020009804A1e1  
; TITLE OF INVENTION: Human G-Proteins  
; FILE REFERENCE: 35800/209285  
; CURRENT APPLICATION NUMBER: US/09/794,257  
; CURRENT FILING DATE: 2001-02-27  
; PRIOR APPLICATION NUMBER: 60/185,606  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 1023  
; TYPE: DNA  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (245)...(886)  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1023)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-794-257-4

Query Match 3.8%; Score 59.6; DB 10; Length 1023;  
Best Local Similarity 72.0%; Pred. No. 7.5e-06;  
Matches 77; Conservative 0; Mismatches 30; Indels 0; Gaps 0;  
QY 1449 TTATTAACAATCATTAACTTGTCACACGGTAAGAAAATACTAGGAATAAAGTAAACGT 1508  
Db 896 TTATTTCCAAAACATGCTCTCTCTACTTGAAGTGAAGAGAAATAAATAGATCTT 955  
QY 1509 CCTGAGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1555  
Db 956 TGTGTNAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1002

RESULT 10  
US-09-925-300-464  
; Sequence 464, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988



;; PRIOR FILING DATE: 2000-03-08  
;; PRIOR APPLICATION NUMBER: 60/124,270  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 1890  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 464  
;; LENGTH: 2431  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-925-300-464

Query Match 3.8%; Score 59.4; DB 10; Length 2431;  
Best Local Similarity 77.4%; Pred. No. 1.4e-05;  
Matches 72; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1463 TTAACCTGTACACGGTAAGAAATACCTTAGGAATAAAGTAAACGCTCCTGAGTAAAAA 1522  
Db 2337 TTGGTTTGTATTCTGTAGCTCAGTGTCTAATAAGTTAAAGATCCTGAAAAA 2396  
QY 1523 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555  
Db 2397 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2429

RESULT 11  
US-10-003-295-1  
;; Sequence 1, Application US/10003295  
;; Patent No. US20020168741A1  
;; GENERAL INFORMATION:  
;; APPLICANT: GAN, Weinlu et al.  
;; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
;; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
;; TITLE OF INVENTION: THEREOF  
;; FILE REFERENCE: CL001183DIV  
;; CURRENT APPLICATION NUMBER: US/10/003,295  
;; CURRENT FILING DATE: 2001-12-06  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 1  
;; LENGTH: 2674  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-10-003-295-1

Query Match 3.8%; Score 59; DB 9; Length 2674;  
Best Local Similarity 55.0%; Pred. No. 1.9e-05;  
Matches 116; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1345 CAGGAGCTGCCTGCACATGCGGATGAGGTGTCGGAGTGGGCGGCCGCGATGGCTCT 1404  
Db 2439 CAGCATCCACACTGCGGCGGAGGATGACGGCCCTGTCTCTGTGTCTGCTGCTGCTGCC 2498  
QY 1405 AGAGTGCCTCTGCTGGCAAGGACAAAGTTATAAGCTATGGGCTTATTACAAATCATT 1464  
Db 2499 AGGGCTTCTCTTCCGGCGAGAAACAATAAACCACTTGTGCCCCTGAAACCACTGAA 2558  
QY 1465 AACTTGTACACGGTAAGAAATACCTTAGGAATAAAGTAAACGCTCCTGAGTAAAAA 1524  
Db 2559 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2618  
QY 1525 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555  
Db 2619 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2649

RESULT 12  
US-09-918-995-30837  
;; Sequence 30837, Application US/09918995  
;; Publication No. US20030073623A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc.  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
;; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

;; FILE REFERENCE: 20411-756  
;; CURRENT APPLICATION NUMBER: US/09/918,995  
;; CURRENT FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: US/09/235,076  
;; PRIOR FILING DATE: 1999-01-20  
;; NUMBER OF SEQ ID NOS: 38054  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 30837  
;; LENGTH: 453  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)...(453)  
;; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-30837

Query Match 3.8%; Score 58.4; DB 9; Length 453;  
Best Local Similarity 48.8%; Pred. No. 1e-05;  
Matches 158; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 1019 AGGTGGAGTGCCTGGTTTCTCTGTCGGATGACAAACCCCTCTATCTGTGGCGGAACACC 1078  
Db 60 ACGGGAGTTTGTAGCCTCTGTGTTCCAGGACACAAACATCAAGCTCTGGGACATCAGGA 119  
QY 1079 AGAACAAAGTGCCTGAGCGCATGACAGGGCACAGAACGTTGTCACGATGTGAATATT 1138  
Db 120 GGAAGGCTGTGTCTTCCGATACAGGGGGCACAGCCAGGCCGTCTCCGGTTCA 179  
QY 1139 CGCCGATGTAAGCTAATTGCGTCTGCTTCAATTGACAAGTCACTGCGTCTGTGGCGAG 1198  
Db 180 GCCCGATGGGAAGTGGTGGCGTGGCCGCGAGATGACCAACCGTGAAGCTCTGGGATC 239  
QY 1199 CCAGCGATGTCAGTACATGTCGTCCTGCTTCCGGGTCATGTGCAGGCTGTTTACACGGTTG 1258  
Db 240 TCACCTCCCGCAAGATGATGTCTGAGTTCCTCTGTCACACGGGCGCTGTCAACGTGGTCG 299  
QY 1259 CCTGGTCCGCGGACTCCCGCTTGATTTTCCGGCAGCAAGACTCAAACTCTAAAAAGTAT 1318  
Db 300 AGTTTACACCCCAACGAGTACCTCCTGGCCTCCGGCAGCTCTGACAGGACAAATCGGCTCT 359  
QY 1319 GGAGTGTCCAGACGAGAGAAACTGG 1342  
Db 360 GGGACCTGGAGAAGTTCAGGTGG 383

RESULT 13  
US-09-997-279-19  
;; Sequence 19, Application US/09997279  
;; Publication No. US20030059781A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Chenault, Ruth A.  
;; APPLICANT: Xu, Jiangchun  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
;; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN AND ENDOMETRIAL CANCER  
;; FILE REFERENCE: 210121.501C1  
;; CURRENT APPLICATION NUMBER: US/09/997,279  
;; CURRENT FILING DATE: 2001-11-28  
;; NUMBER OF SEQ ID NOS: 230  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 19  
;; LENGTH: 1602  
;; TYPE: DNA  
;; ORGANISM: Homo sapien  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)...(1602)  
;; OTHER INFORMATION: n = A,T,C or G  
US-09-997-279-19

Query Match 3.7%; Score 57.6; DB 9; Length 1602;  
Best Local Similarity 65.1%; Pred. No. 3.3e-05;  
Matches 84; Conservative 0; Mismatches 45; Indels 0; Gaps 0;



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 04:48:24 ; Search time 2084 Seconds  
(without alignments)  
12084.444 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattcccaaaaatgcagga.....aaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	735.6	47.3	790	9 AI518473	AI518473 LD38001.5
2	597.8	38.4	637	13 BI637470	BI637470 SD19506.5
3	591.8	38.1	725	12 BF505998	BF505998 AT08344.5
4	578.2	37.2	622	13 BI632742	BI632742 SD26605.5
5	549.2	35.3	606	13 BI232464	BI232464 RE28535.5
6	548.8	35.3	626	13 BI639652	BI639652 SD22291.5

7	488.2	31.4	527	9	AI520108	AI520108 LD40082.5
8	485.2	31.2	552	13	BI635777	BI635777 SD17257.5
9	476.6	30.6	1784	11	BC035421	BC035421 Mus muscu
10	455.8	29.3	510	13	BI635228	BI635228 SD16515.5
11	424	27.3	448	9	AA735870	AA735870 GM10004.5
12	346	22.3	688	13	BM579586	BM579586 170006872
13	331.6	21.3	975	9	AL520339	AL520339 AL520339
14	314	20.2	925	9	AL526575	AL526575 AL526575
15	301.4	19.4	1114	9	AL530778	AL530778 AL530778
16	296.4	19.1	861	9	AL558947	AL558947 AL558947
17	290.4	18.7	851	9	AL558453	AL558453 AL558453
18	289.4	18.6	861	9	AL519702	AL519702 AL519702
19	284.2	18.3	997	14	BQ070427	BQ070427 AGENCOURT
20	282	18.1	927	12	BG470097	BG470097 602533586
21	278.8	17.9	715	13	BM579239	BM579239 170006872
22	278.8	17.9	939	14	BQ956834	BQ956834 AGENCOURT
23	276.6	17.8	826	9	AL526922	AL526922 AL526922
24	275.2	17.7	822	9	AL550330	AL550330 AL550330
25	268.8	17.3	1103	13	BM478420	BM478420 AGENCOURT
26	265.2	17.1	1006	13	BM474111	BM474111 AGENCOURT
27	263.6	17.0	816	9	AL527738	AL527738 AL527738
28	263	16.9	573	13	BM654928	BM654928 170006873
29	262.4	16.9	324	13	BI232484	BI232484 RE28559.5
30	256.6	16.5	921	13	BI833055	BI833055 603090947
31	251.4	16.2	912	9	AL522417	AL522417 AL522417
32	250.2	16.1	903	12	BG481047	BG481047 602529007
33	245.4	15.8	673	13	BM654934	BM654934 170006873
34	242.6	15.6	678	14	BQ257262	BQ257262 NISC.ko12
35	242.6	15.6	776	12	BG323761	BG323761 602421865
36	242	15.6	846	13	BI116063	BI116063 602866459
37	241.2	15.5	1080	13	BM467429	BM467429 AGENCOURT
38	239.4	15.4	756	12	BG386139	BG386139 602455319
39	239	15.4	897	14	BQ727393	BQ727393 AGENCOURT
40	238.4	15.3	697	13	BI855975	BI855975 603383349
41	237.4	15.3	843	12	BE792381	BE792381 601585118
42	236.2	15.2	814	13	BI772615	BI772615 603057607
43	233.6	15.0	759	12	BF536680	BF536680 602048656
44	232.6	15.0	274	9	AA696993	AA696993 GM08866.5
45	229.8	14.8	668	14	BM849233	BM849233 K-EST0129

ALIGNMENTS

RESULT 1  
AI518473  
LOCUS  
DEFINITION

AI518473 790 bp mRNA linear EST 23-APR-2001  
LD38001.5prime LD Drosophila melanogaster embryo pot2 Drosophila  
melanogaster cDNA clone LD38001 5 similar to Nle: FBan0002863  
'signal transduction' located on: 2L 21C6-21C6;: 04/10/2001, mRNA  
sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AI518473  
AI518473.2 GI:13769352  
EST.  
fruit fly.  
Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
AUTHORS

1 (bases 1 to 790)  
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,  
Lewis,S. and Rubin,G.M.

TITLE  
JOURNAL  
COMMENT

BDGP/HMI Drosophila EST Project  
Unpublished (2001)  
On Mar 16, 1999 this sequence version replaced gi:4424327.  
Other ESTs: LD38001.3prime  
Contact: Stapleton, M.  
BDGP

Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)  
hit genomic AE003589: arm:2L [301580,604052]

estimated-cyto:21B7-21C7; 04/10/2001  
Plate: LD.380 row: A column: 1  
High quality sequence stop: 626  
POLYA-No.

FEATURES  
source

Location/Qualifiers

1..790  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="LD38001"  
/clone\_lib="LD Drosophila melanogaster embryo pot2"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="XLI Blue"  
/note="Organ: embryo; Vector: pot2; Site\_1: EcoRI; Site\_2:  
XhoI; Sized fractionated cDNAs were directly ligated into  
pot2."

BASE COUNT 199 a 198 c 247 g 146 t  
ORIGIN

Query Match 47.3%; Score 735.6; DB 9; Length 790;  
Best Local Similarity 98.8%; Pred. No. 8.1e-110;  
Matches 763; Conservative 0; Mismatches 4; Indels 5; Gaps 2;

QY 7 CAAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACCCACATACCATACAGGCGCGC 66  
Db 19 CAGAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACCCACATACCATACAGGCGCGC 78  
QY 67 CTCGTTT---ACACGGGCGAGGAAGCCGCCGCCCAATCGACCTGCCGGCAGGAATCACT 123  
Db 79 CTCGTTTCGGACACGGCGAGGAAGCGGCCGCCCAATCGACCTGCCGGCAGGAATCACT 138  
QY 124 ACCCAGCAATGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT 183  
Db 139 ACCCAGCAATGGGTCTGATTTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT 198  
QY 184 TTGTTTTTCGTGGCGGAGGATGATCAAGAAGAGCCCTGGAGGACACGTTGGACTTGGCG 243  
Db 199 TTGTTTTTCGTGGCGGAGGATGATCAAGAAGAGCCCTGGAGGACACGTTGGACTTGGCG 258  
QY 244 TCAGTGGACACCGAAACGTTGATCGATATTTGTATCAGCCACACAGGCGGTTTTCAAAGTG 303  
Db 259 TCAGTGGACACCGAAACGTTGATCGATATTTGTATCAGCCACACAGGCGGTTTTCAAAGTG 318  
QY 304 CGCCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 363  
Db 319 CGCCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 378  
QY 364 AATTTACGCCCGGATGGTCTCATCTGCCAGTGGAAAGTGCGGACAC--CACAGTGGCAT 421  
Db 379 AATTTACGCCCGGATGGTCTCATCTGCCAGTGGAAAGTGCGGACACACACAGTGGCAT 438  
QY 422 TGTGGGATCTTAACACAGACACCCGACTTCACCTGCACAGGTATATAAGCAGTGGGTTTC 481  
Db 439 TGTGGGATCTTAACACAGACACCCGACTTCACCTGCACAGGTATATAAGCAGTGGGTTTC 498  
QY 482 TGTGCGTATCCTGGGCTCCGGATGGCAACGGTTTGGCCAGCGGTTGCAAAAGCGGGCTCTA 541  
Db 499 TGTGCGTATCCTGGGCTCCGGATGGCAACGGTTTGGCCAGCGGTTGCAAAAGCGGGCTCTA 558  
QY 542 TAATCATCTGGGACCGGAGACGGGTACGACAGAGGGGCGACCCCTTGAGTGGGCACAAGA 601  
Db 559 TAATCATCTGGGACCGGAGACGGGTACGACAGAGGGGCGACCCCTTGAGTGGGCACAAGA 618  
QY 602 AACACATCAACTGCCCTCGCCTGGGACCGGTATCATCGCGATCCGGAGTGCAGAAACTTG 661  
Db 619 AACACATCAACTGCCCTCGCCTGGGACCGGTATCATCGCGATCCGGAGTGCAGAAACTTG 678  
QY 662 CTTCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAAATTTGGCCAGTGCCTTA 721  
Db 679 CTTCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTGAATTTGGCCAGTGCCTTA 738  
QY 722 TGAACATTCGCCGGACACACAAATGCTGTGACACAGTGAATGGGTTGGAGC 773  
|||||

Db 739 TGAACATTCGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGC 790

RESULT 2  
BI637470

LOCUS

DEFINITION

BI637470 637 bp mRNA linear EST 10-SEP-2001  
SD19506.5prime SD Drosophila melanogaster Schneider L2 cell culture  
pot2 Drosophila melanogaster cDNA clone SD19506 5 similar to Nle:  
FBan0002863 GO:[signal transduction (GO:0004871)] located on: 2L  
21C6-21C6;: 05/19/2001, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BDGP  
Contact: Stapleton, M.  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu  
hit genomic AB003589: arm:2L [301580,604052]  
estimated-cyto:21B7-21C7: 05/19/2001  
Plate: SD.195 row: A column: 6  
High quality sequence stop: 614.  
Location/Qualifiers

FEATURES

source

1..637  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="SD19506"  
/clone\_lib="SD Drosophila melanogaster Schneider L2 cell  
culture pot2"  
/lab\_host="DH5-alpha"  
/note="Vector: pot2; Site\_1: EcoRI; Site\_2: XhoI; Sized  
fractionated cDNAs were directly ligated into pot2.  
Plasmid cDNA library."  
BASE COUNT 164 a 164 c 194 g 115 t  
ORIGIN

Query Match 38.4%; Score 597.8; DB 13; Length 637;  
Best Local Similarity 99.2%; Pred. No. 1.7e-87;  
Matches 612; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 7 CAAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACCCACATACCATACAGGCGCGC 66  
Db 21 CAGAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACCCACATACCATACAGGCGCGC 80  
QY 67 CTCGTTT---ACACGGGCGAGGAAGCCGCCGCCCAATCGACCTGCCGGCAGGAATCACT 123  
Db 81 CTCGTTTCGGACACGGGCGAGGAAGCCGCCGCCCAATCGACCTGCCGGCAGGAATCACT 140  
QY 124 ACCCAGCAATGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT 183  
Db 141 ACCCAGCAATGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT 200  
QY 184 TTGTTTTTCGTGGCGGAGGATGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 243  
Db 201 TTGTTTTTCGTGGCGGAGGATGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 260  
QY 244 TCAGTGGACACCGAAACGTTGATCGATATTTGTATCAGCCACACAGGCGGTTTTCAAAGTG 303  
Db 261 TCAGTGGACACCGAAACGTTGATCGATATTTGTATCAGCCACACAGGCGGTTTTCAAAGTG 320  
QY 304 CGCCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 363  
|||||



Db 321 CGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCGGAGGCTGTGGTTTCGCTG 380

QY 364 AATTTACGCCCGGATGGTGCTCATCTCGCCAGTGGAGTGGCGACACACACAGTGGGATTG 423

Db 381 AATTTACGCCCGGATGGTGCTCATCTCGCCAGTGGAGTGGCGACACACACAGTGGGATTG 440

QY 424 TGGGATCTTAACACAGACACACCGGCTCACCTGCACAGGTGCATAAGCAGTGGGTTCTG 483

Db 441 TGGGATCTTAACACAGACACACCGGCTCACCTGCACAGGTGCATAAGCAGTGGGTTCTG 500

QY 484 TGGGATCTTCCGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCCAAAGCGGGCTCTATA 543

Db 501 TGGGATCTTCCGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCCAAAGCGGGCTCTATA 560

QY 544 ATCATCTGGGACCCCGGAGACGGGTGCAGAGAGGGGGACCCCTTGAGTGGGCACAGAAAA 603

Db 561 ATCATCTGGGACCCCGGAGACGGGTGCAGAGAGGGGGACCCATTGAGTGGGCACAGAAAA 620

QY 604 CACATCAACTGCCTCGC 620

Db 621 CACATCAACTGCCTCGC 637

RESULT 3

BF505998

LOCUS

DEFINITION

AT08344.5prime AT Drosophila melanogaster adult testes pOTB7

Drosophila melanogaster cDNA clone AT08344 5 similar to Nle:

FBan0002863 'signal transduction' located on: 2L 21C6-21C6;;

04/07/2001, mRNA sequence.

ACCESSION

BF505998

VERSION

BF505998.2 GI:13688580

KEYWORDS

EST.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

1 (bases 1 to 725)

Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman

,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan

,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,

Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J.,

Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E.,

Celniker,S. and Rubin,G.M.

BDGP/HMI AT Drosophila EST Project

Unpublished (2000)

On Dec 6, 2000 this sequence version replaced gi:11589299.

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: [http://www.fruitfly.org/EST\\_est@fruitfly.berkeley.edu](http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu)

hit genomic AE003589: arm:2L [301580,604052]

estimated-cyto:21B7-21C7; 04/07/2001

Plate: At.83 row: D column: 8

High quality sequence stop: 661.

FEATURES

source

1..725

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="AT08344"

/clone\_lib="AT Drosophila melanogaster adult testes pOTB7"

/sex="male"

/dev\_stage="0-3 day old Ore-R males"

/lab\_host="Plates At.10-AT.120: DH5-alpha. Plates

AT.121-AT.319: DH5-alpha Tona"

/note="Organ: ADULT testes; Vector: pOTB7; Site\_1: EcORI;

Site\_2: XhoI; The mRNA for the testis library was made

from testes and seminal vesicles hand dissected from 0-3

day old Ore-R males. RNA kindly provided by the lab of

Margaret Fuller. Sized fractionated cDNAs were directly

BASE COUNT 187 a 182 c 215 g 141 t

ORIGIN

Query Match 38.1%; Score 591.8; DB 12; Length 725;

Best Local Similarity 99.2%; Pred. No. 1.5e-86;

Matches 606; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 7 CAAAAAATGCAGGAGACGGACAGGAGCAAGAGGCCACGCACATACGATACAGGCGCGC 66

Db 115 CAGAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCACATACGATACAGGCGCGC 174

QY 67 CTCGTTT---ACACGGGGGAGGAAGCCGCCGCCCAATCGACCTGCCGGCAGGAATCACT 123

Db 175 CTCGTTTCGGACACGGGGGAGGAAGCCGCCGCCCAATCGACCTGCCGGCAGGAATCACT 234

QY 124 ACCCAGCAATTGGGACTGATTTGCAACCGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 183

Db 235 ACCCAGCAATTGGGACTGATTTGCAACCGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 294

QY 184 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAGAGCCCTGGAGGACACGTTGGACTTGGCG 243

Db 295 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAGAGCCCTGGAGGACACGTTGGACTTGGCG 354

QY 244 TCAGTGGACACCGAAAAACGTGATCGATATTGTGATATCAGCCACAGCGGGTTTCAAAGTG 303

Db 355 TCACTGGACACCGAAAAACGTGATCGATATTGTGATATCAGCCACAGCGGGTTTCAAAGTG 414

QY 304 CGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCAGGCTGTGGTTTCGCTG 363

Db 415 CGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCAGGCTGTGGTTTCGCTG 474

QY 364 AATTCAGCCCGGATGGTGCTCATCTGCCAGTGGGAAGTGGCGACACACAGTGGGATTG 423

Db 475 AATTCAGCCCGGATGGTGCTCATCTGCCAGTGGGAAGTGGCGACACACAGTGGGATTG 534

QY 424 TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCAGTGGTTCG 483

Db 535 TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCAGTGGTTCG 594

QY 484 TGCATATCTGGGCTCGGATGGCAAAACGGTTGGCCAGCGGTTGCCAAAGCGGGCTCTATA 543

Db 595 TGCATATCTGGGCTCGGATGGCAAAACGGTTGGCCAGCGGTTGCCAAAGCGGGCTCTATA 654

QY 544 ATCATCTGGGACCCCGGAGACGGGTGCAGAGAGGGGGACCCCTTGAGTGGGCACAGAAA 603

Db 655 ATCATCTGGGACCCCGGAGACGGGTGCAGAGAGGGGGACCCATTGAGTGGGCACAGAAA 714

QY 604 CACATCAACTG 614

Db 715 CACATCAACTG 725

RESULT 4

BI632742

LOCUS

DEFINITION

SD26605.5prime SD Drosophila melanogaster Schneider L2 cell culture

pOT2 Drosophila melanogaster cDNA clone SD26605 5 similar to Nle:

FBan0002863 GO:[signal transduction (GO:0004871)] located on: 2L

21C6-21C6;; 05/23/2001, mRNA sequence.

ACCESSION

BI632742

VERSION

BI632742.1 GI:15534952

KEYWORDS

EST.

SOURCE

fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

1 (bases 1 to 622)

Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,

Lewis,S. and Rubin,G.M.

BDGP/HMI Drosophila EST Project

Unpublished (2001)

COMMENT

Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu>  
hit genomic AE003589: arm:2L [301580,604052]  
estimated-cyto:21B7-21C7: 05/23/2001  
Plate: SD.266 row: A column: 5  
High quality sequence stop: 499.  
Location/Qualifiers  
1. .622  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="SD26605"  
/clone\_lib="SD Drosophila melanogaster Schneider L2 cell culture pot2"  
/lab\_host="DH5-alpha"  
/note="Vector: pot2; Site\_1: EcoRI; Site\_2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library."  
BASE COUNT 162 a 159 c 188 g 113 t

FEATURES

source

Query Match 37.2%; Score 578.2; DB 13; Length 622;  
Best Local Similarity 98.9%; Pred. No. 2.6e-84;  
Matches 604; Conservative 0; Mismatches 3; Indels 4; Gaps 2;  
QY 7 CAAAAATGCAGGACGACGGACAGAGAGGCCACATACGATACAGGCGGC 66  
|| |||||  
Db 13 CAGAAATGCAGGACGACGGACAGAGAGGCCACATACGATACAGGCGGC 72  
QY 67 CTCGTTT--ACACGGCGGAGAACCGGCCCAATCGACCTGCCGGCAGGAATCACT 123  
|| |||||  
Db 73 CTCGTTTCGGACACGGCGGAGAACCGGCCCAATCGACCTGCCGGCAGGAATCACT 132  
QY 124 ACCCAGCAATTGGGACTGATTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT 183  
|| |||||  
Db 133 ACCAGCAATTGGGACTGATTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT 192  
QY 184 TTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCCTGGAGGACAGCTTGGACTGGCG 243  
|| |||||  
Db 193 TTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCCTGGAGGACAGCTTGGACTGGCG 252  
QY 244 TCAGTGGACACGAAACGTGATCGATATGTGTATCAGCCACAGGCGGTTTCAAAGTG 303  
|| |||||  
Db 253 TCAGTGGACACGAAACGTGATCGATATGTGTATCAGCCACAGGCGGTTTCAAAGTG 312  
QY 304 CGCCAGTGACAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 363  
|| |||||  
Db 313 CGCCAGTGACAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 372  
QY 364 AATTTCAGCCCGGATGGTGTCTCATCTCGCCAGTGGAGTGGCGACACACAGTGGGATTG 423  
|| |||||  
Db 373 AATTTCAGCCCGGATGGTGTCTCATCTCGCCAGTGGAGTGGCGACACACAGTGGGATTG 432  
QY 424 TGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTTCATAGCAGTGGTCTG 483  
|| |||||  
Db 433 TGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTTCATAGCAGTGGTCTG 492  
QY 484 TGGTATCTTGGCTCCGGATGGCAACGGTTGGCCAGCGTTGCAAGCGGGCTCTATA 543  
|| |||||  
Db 493 TGGTATCTTGGCTCCGGATGGCAACGGTTGGCCAGCGTTGCAAGCGGGCTCTATA 552  
QY 544 ATCATCTGGACCCGGAGACGGGTACAGAAAGGGGCCACCCCTTGACTGGGCACAGAAA 603  
|| |||||  
Db 553 ATCATCTGGACCCGGAGAC--GGTCAGCAGAAAGGGGCCACCATTTGAGTGGGCACAGAAA 611  
QY 604 CACATCAACTG 614  
|| |||||  
Db 612 CACATCAACTG 622

RESULT 5

BI232464

LOCUS

DEFINITION

BI232464 606 bp mRNA linear EST 11-JUL-2001  
RE28535.5prime RE Drosophila melanogaster normalized Embryo pFlc-1  
Drosophila melanogaster cDNA clone RE28535 5 similar to Nle;  
FBan0002863 'signal transduction' located on: 2L 21C6-21C6;;  
04/12/2001, mRNA sequence.  
BI232464  
BI232464.1 GI:14700036

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

1 (bases 1 to 606)  
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson,  
J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George,  
R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Misra, S.,  
Mungall, C. J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,  
Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celniker, S. and Rubin,  
G. M.  
BDGP/HMI RE Drosophila EST Project  
Unpublished (2001)  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu>  
hit genomic AE003589: arm:2L [301580,604052]  
estimated-cyto:21B7-21C7: 04/12/2001  
Plate: RE.285 row: C column: 11  
High quality sequence stop: 546.  
Location/Qualifiers  
1. .606

TITLE

JOURNAL

COMMENT

FEATURES

source

/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="RE28535"  
/clone\_lib="RE Drosophila melanogaster normalized Embryo pFlc-1"  
/sex="male and female"  
/dev\_stage="0-24 hours mixed stage embryonic"  
/lab\_host="DH5-alpha Tona"  
/note="Organ: embryo; Vector: pFlc1; Site\_1: XhoI; Site\_2:  
BamHI; Library was kindly generated by Piero Carninci at  
the RIKEN. The library was normalized and excised using  
Cre recombinase. Plasmid cDNA library."  
BASE COUNT 150 a 153 c 185 g 118 t

ORIGIN

Query Match 35.3%; Score 549.2; DB 13; Length 606;  
Best Local Similarity 98.9%; Pred. No. 1.3e-79;  
Matches 564; Conservative 0; Mismatches 3; Indels 3; Gaps 1;  
QY 7 CAAAAATGCAGGACGACGACGAGCAAGAGCCACGCCACATACGATACAGGCGGC 66  
|| |||||  
Db 37 CAGAAATGCAGGACGACGACGAGCAAGAGCCACGCCACATACGATACAGGCGGC 96  
QY 67 CTCGTTT--ACACGGCGGAGGAGCCGCCCAATCGACCTGCCGGCAGGAATCACT 123  
|| |||||  
Db 97 CTCGTTTCGGACACGGCGGAGGAGCGGCCCAATCGACCTGCCGGCAGGAATCACT 156  
QY 124 ACCCAGCAATTGGGACTGATTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT 183  
|| |||||  
Db 157 ACCCAGCAATTGGGCTGATTGCAACGCGCTGCTGAAAAACGAGGAGCCACTCCATAT 216  
QY 184 TTGTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTTGGACTTGGCG 243  
|| |||||  
Db 217 TTGTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTTGGACTTGGCG 276  
QY 244 TCAGTGGACACCGAAACGTCGATGATATTTGTCATCAGCCACAGCGGTTTCAAAGTG 303  
|| |||||

Db	277	TCAGTGGACACCGAAACGTTGATCGATATGTGTATCAGCCACAGCGGTTTTCAAAGTG	336
QY	304	CGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG	363
Db	337	CGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG	396
QY	364	AATTTACGCCCGGATGGTGTCTATCTCGCAGTGGAAAGTGGCGACACACAGTGGCATTG	423
Db	397	AATTTACGCCCGGATGGTGTCTATCTCGCAGTGGAAAGTGGCGACACACAGTGGCATTG	456
QY	424	TGGATCTTAACACAGAGACACCGCACCTTCACCTGCACAGGTCATAAGCAGTGGGTTCTG	483
Db	457	TGGATCTTAACACAGAGACACCGCACCTTCACCTGCACAGGTCATAAGCAGTGGGTTCTG	516
QY	484	TGGTATCTCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCCAAGCGGGCTCTATA	543
Db	517	TGCGTATCTCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCCAAGCGGGCTCTATA	576
QY	544	ATCATCTGGGACCCGGAGACGGGTCAGCAG	573
Db	577	ATCATCTGGGACCCGGAGACGGGTCAGCAG	606
RESULT 6			
BI639652			
LOCUS			
DEFINITION			
SD22291.5prime SD Drosophila melanogaster Schneider L2 cell culture			
pot2 Drosophila melanogaster cDNA clone SD22291 5 similar to Nle:			
FBan0002863 GO:[signal transduction (GO:0004871)] located on: 2L			
21C6-21C6:: 05/19/2001, mRNA sequence.			
ACCESSION			
BI639652			
VERSION			
BI639652.1 GI:15541862			
KEYWORDS			
EST.			
SOURCE			
fruit fly.			
ORGANISM			
Drosophila melanogaster			
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE			
1 (bases 1 to 626)			
AUTHORS			
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,			
Lewis,S. and Rubin,G.M.			
TITLE			
BDGP/HMI Drosophila EST Project			
JOURNAL			
Unpublished (2001)			
COMMENT			
Contact: Stapleton, M.			
BDGP			
Lawrence Berkeley National Lab			
One Cyclotron Rd, Berkeley, CA 94720, USA			
Fax: 510 486 6798			
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu			
hit genomic AE003589: arm:2L [301580,604052]			
estimated-cyto:21B7-21C7: 05/19/2001			
Plate: SD.222 row: H column: 7			
High quality sequence stop: 544.			
FEATURES			
source			
1. .626			
/organism="Drosophila melanogaster"			
/db_xref="taxon:7227"			
/clone="SD22291"			
/clone_lib="SD Drosophila melanogaster Schneider L2 cell			
culture pot2"			
/lab_host="DH5-alpha"			
/note="Vector: pot2; Site_1: EcoRI; Site_2: Xho1; Sized			
fractionated cDNAs were directly ligated into pot2.			
Plasmid cDNA library."			
BASE COUNT			
155 a 158 c 188 g 124 t 1 others			
ORIGIN			
Query Match			
Best Local Similarity 35.3%; Score 548.8; DB 13; Length 626;			
Matches 563; Conservative 0; Mismatches 3; Indels 3; Gaps 1;			
QY			
7 CAAAAATGCAGGACGGACACGGAGCAAGAGCCACGCCACATACGATACAGGCGGC			
66			

Db	58	CAGAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGGC	117
QY	67	CTCGTTT---ACACGGCGGAGGAAGCGCGCCGCCCAANTCGACCTGCCGCGAGGAATCACT	123
Db	118	CTCGTTTCGGACACGGCGGAGGAAGCGCGCCGCCAATCGACCTGCCGCGAGGAATCACT	177
QY	124	ACCCAGCAATTGGGACTGATTTCGAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT	183
Db	178	ACCCAGCAATTGGGACTGATTTCGAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT	237
QY	184	TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGACGCTGGAGGACACGTTGGACTTGGCG	243
Db	238	TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGACGCTGGAGGACACGTTGGACTTGGCG	297
QY	244	TCAGTGGACACCCGAAAAACGTCGATATTCGATATTCGATCAGCCACAGCGGTTTCAAAAGTG	303
Db	298	TCAGTGGACACCCGAAAAACGTCGATATTCGATATTCGATCAGCCACAGCGGTTTCAAAAGTG	357
QY	304	CGCCAGTGACAAGATGCACGAGTTCATCCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG	363
Db	358	CGCCAGTGACAAGATGCACGAGTTCATCCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG	417
QY	364	AATTCAGCCCCGGATGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGGGATTG	423
Db	418	AATTCAGCCCCGGATGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGGGATTG	477
QY	424	TGGGATCTTAACACAGAGACACCGGCACTTCACCTGCACAGGTCATAAGCAGTGGGTTCTG	483
Db	478	TGGGATCTTAACACAGAGACACCGGCACTTCACCTGCACAGGTCATAAGCAGTGGGTTCTG	537
QY	484	TGCGTATCTCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAAAGCGGCTCTATA	543
Db	538	TGCGTATCTCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAAAGCGGCTCTATT	597
QY	544	ATCATCTGGGACCCCGGAGACGGGTCAGCA	572
Db	598	ATCATCTGGGACCCCGGAGACGGGTCAGCA	626
RESULT 7			
AI520108			
LOCUS			
DEFINITION			
LD40082.5prime LD Drosophila melanogaster embryo pot2 Drosophila			
melanogaster cDNA clone LD40082 5prime, mRNA sequence.			
ACCESSION			
AI520108			
VERSION			
AI520108.1 GI:4425962			
KEYWORDS			
EST.			
SOURCE			
fruit fly.			
ORGANISM			
Drosophila melanogaster			
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE			
1 (bases 1 to 527)			
AUTHORS			
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,			
Lewis,S. and Rubin,G.M.			
TITLE			
BDGP/HMI Drosophila EST Project			
JOURNAL			
Unpublished (2001)			
COMMENT			
Contact: Stapleton, M.			
BDGP			
Lawrence Berkeley National Lab			
One Cyclotron Rd, Berkeley, CA 94720, USA			
Fax: 510 486 6798			
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu			
hit genomic sequence AC004573			
Plate: 400 row: G column: 10			
High quality sequence stop: 506.			
FEATURES			
source			
1. .527			
/organism="Drosophila melanogaster"			
/db_xref="taxon:7227"			
/clone="LD40082"			
/clone_lib="LD Drosophila melanogaster embryo pot2"			
/sex="male and female"			







AUTHORS
TITLE
JOURNAL

Strausberg, R.  
Direct Submission  
Submitted (31-JUL-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

NIH-MGC Project URL: <http://mac.nci.nih.gov>

Contact: MGC help desk

contact: noc\_help@ccn  
Email: ccapps-remail.nih.gov

Email: cyprus1email@nfl.gov  
 Tissue procurement: Lothar Hennelhausen Ph.D., Robin Humphreys

ISSUE FOCUSED: Journal Technology F.I.D.; RNA Library Preparation: Life Technologies, Inc.

CDNA Library Preparation: LIFE Technologies, Inc.  
CDNA Library Arrayed by: The T M A G F Consortium (I.I.N.T.)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome

DNA sequencing by  
sequencing center

Sequencing Center  
Center code: PCM-UCSC

Center code: BCM-HGSC  
Web site: <http://www.bcm.tmc.edu/cdna/>

Web Site: <http://www.ngs>  
Contact: [web@ngs](mailto:web@ngs) tmo adj:

Contact: amge@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowls, C.R.,

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 30 Row: a Column: 14  
 This clone has the following problem: no 5' EST match.

## FEATURES

sources

```

1. 1784
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:2811006"
/tissue_type="Mammary tu
old, gross tissue."
/clone_lib="NCI_CGAP_Mamm
/lab_host="DH10B"
/note="Vector: pCMV-SPOR
521c 539 q 32

```

BASE COUNT	397 a	521 c	539 q	327 t
BASE COUNT	397 a	521 c	539 q	327 t

## ORIGIN

Query Match	30.6%;	Score 476.6;	DB 11;	Length 1784;
Best Local Similarity	59.2%;	Pred. No. 4.6e-68;		
Matches 850;	Conservative	0;	Mismatches 579;	Indels
				Gaps 2;

Qy	18	GGAGACGGACACGAGCAAGAGGCCACGCCACATACCATACAGGC	GCGCCTCGTTTACAC	77
Dd	45	GGTGGAAGAGCGCGCGCACGGCAGCGTCGTACAGTCCAGGATGAGGG	104	
Qy	78	GGCGGAGGAAGCGCGGCCGCCAATCGACTGCCGGCAGGAATCACTACCAGCAATTGGG	137	
Dd	105	CGBTGAGCTGCTGGGGCTCCCGTTCGACGTGCCCGTGGACATCACGCCGGACAAGCTGCA	164	
Qy	138	ACTGATTTGCAACGCGCTGCTGAATAAACGAGGAAGCCAATTCATATTGTTTTTCGTGGG	197	
Dd	165	GCTGGTGTCAACGCGCTGCTGGCCCCAGGAGGAGCCCTGCCCTGGCGTTCATGTCCA	224	
Qy	198	CGAGGATGAGATCAAAGAAGACCTGGAGGACACGTTGSACTTGGCGTCAGTGGACACCGA	257	
Dd	225	CGATGCTGAGATTGTCTCCCTGGSGAAGACGCTGAGTCCCAGTCTGTGGAGACAGA	284	
Qy	258	AACGCTGATCGATATTGTGTATCAGCCACAGCGGGTTTCAAAGTCGCCCAAGTAGACAAG	317	
Dd	285	GAAAATCGTCGATATCATCTACCAGGCCCAAGCTGTCTCAGGGTCCGGGCTGTGACGCG	344	
Qy	318	ATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGGTTCGCTGAATTCAGCCCGGA	377	
Dd	345	GTGCACCACTCATTTGAGGGTCA CAGCGAAAGCTGT CATTTCTGTGGCCTTCAGCCCCAC	404	
Qy	378	TGGTGCTCATCTCGCCAGTGGGAAGTGGGACACCACTGCGATTGTGGGATCTTAACAC	437	
Dd	405	AGAAAAATACCTGGCCAGTGGTCCGGAGACACCACTGTGCGCTTCTGGGATCTCAGCAC	464	
Qy	438	AGAGACACCGACTTCA CCTGCACAGGTCA TAGCAGTGGGTTCTGTGCGTATCTCTGGGC	497	



|||||  
Db 301 ACCAGTTCCATGCCGGGACAGCCGAGGCTGTGGTTTCGCTGAATTCAGCCCGGATGGT 360  
382 GCTCATCTCGCCAGTGGGAAGTGGCGACACACAGTGCAGATTGTGGGATCTTAAACACAGAG 441  
Db 361 GCTCATCTCGCCAGTGGGAAGTGGCGACACACAGTGCAGATTGTGGGATCTTAAACACAGAG 420  
QY 442 ACACCGCACTTCACCTGCACAGGTCATA 469  
Db 421 ACACCGCACTTCACCTGCACAGGTCATA 448

RESULT 12  
BM579586  
LOCUS  
DEFINITION 17000687239580 A.Gam.ad.cdna.blood1 Anopheles gambiae cDNA clone  
19600449699487 5', mRNA sequence.  
ACCESSION BM579586  
VERSION BM579586.1 GI:18868053  
KEYWORDS EST.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
Anopheles.  
REFERENCE 1 (bases 1 to 688)  
AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab  
R., Collins,F.H., Venter,J.C. and Hoffman,S.L.  
TITLE Celera Anopheles gambiae EST project  
JOURNAL Unpublished (2002)  
COMMENT Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 2404534580  
Email: Holtra@celera.com  
Plate: NU01004AB3 row: D column: 01  
Seq primer: M13 Reverse.

FEATURES  
source  
1. .688  
/organism="Anopheles gambiae"  
/strain="RSP-ST (Reduced susc. to Permethrin - std.  
chromosome)"  
/db\_xref="taxon:7165"  
/clone="19600449699487"  
/clone\_lib="A.Gam.ad.cdna.blood1"  
/dev\_stage="Adult"  
/lab\_host="DH10b"  
/note="Vector: pSport1; Site\_1: SalI; Site\_2: NotI; Whole  
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24  
hours after human blood feeding. cDNA inserts >500 bp  
cloned directionally into pSport 1. Not 1 site is 3'.  
Clones available through the Malaria Research and  
Reference Reagent Resource Center (www.malaria.mr4.org)"  
BASE COUNT 136 a 178 c 240 g 134 t

Query Match 22.3%; Score 346; DB 13; Length 688;  
Best Local Similarity 71.0%; Pred. No. 9e-47;  
Matches 473; Conservative 0; Mismatches 190; Indels 3; Gaps 1;  
QY 793 TCCAAGATCGACAGTGAAGATGTGGCGACGAGCTGATGGAATCTGTGCGCGAGCTTC 852  
Db 14 TCCCGGATCGGACGGTCAAGATGTGGCGAGGACGGCGTACTGTGCAAAACGTTTC 73  
QY 853 TCTGGCCCAAGCTCACTGGGTAACAACATTCGGCTGAGCAGCAGATTACGTCCTGCGCACT 912  
Db 74 ACGGTCACGCGCACTGGGTGAATAATTGGCGCTCACACGAGCTAGCTGCTGCGGACG 133  
QY 913 GGTCCATTCCATCCGGTGAAGATCGCTCCAAGAGCCAC---CTCAGTTTGAGCACTGAG 969  
Db 134 GGCCCGTTTCACCCGGTGATGGACAAATTAAAGATGTACGGGGACGGCGCGGATAGGAC 193

QY 970 GAATTCAGGAATCTGCTTTGAAGCGCTACCAGGCGGTGTCCCTGACGAGTGGAGTCG 1029  
Db 194 GCCCTCAAGCAGCTGGCGCTCGAGCGGTACGAGGCTGTCCCGGACGGGTGGAATCG 253  
QY 1030 CTGGTTTCTCTGTTCCGATGACAACACCTCTATCTGTGGCGGAACAACAGAAAGTGC 1089  
Db 254 TTCGTCTCTGTTTCGGACGACTTTACGCTGTATCTGTGGCGCGCAACAGAACTTC 313  
QY 1090 GTTGAGCGCATGACAGGCGCACAGACGTTGTCACGATGTGAATATTCGCCGGATGTA 1149  
Db 314 GTGACGCGCATGACCGGCGCACAGATGTGGTGAAGATGTGAAGTACTCGCCGACGTG 373  
QY 1150 AAGCTAATTCGCTCTGCTTTCATTTGACAAAGTCAGTCGCTCTGTGGCGAGCCAGCATGGT 1209  
Db 374 AAGTTTATTCGCTCCGCTTCGTTTCGACAAATCGATCCGGCTGTGGCGGCGCGCATGGG 433  
QY 1210 CAGTACATGGCCACCTTCCGGGGTCATGTGCAGGCTGTTCACACGTTGCCCTGTCGCGG 1269  
Db 434 CGTTTCATTTGTGCTTCCGCGGTACGCTGCTGGCGCTCTACACGCTGAGCTGCTCGGCC 493  
QY 1270 GACTCCCGCTGATTTGTTCCGGCAGCAAAAGACTCAACTCTAAAAGTATGGAGTGCAG 1329  
Db 494 GATTCGCGCTAGTGTCTAGCGGCGCAGCAAGGACTCAACCTCAAGGTGTGGAGTGTGAAG 553  
QY 1330 ACGAAGAACTGGCACAGGAGCTGCCTGGACATCGGATGAGGTGTTCCGAGTGGACTGG 1389  
Db 554 GAGCGCAAGCTGGTGCAGGAGCTGCCCGGACACGCGCAGAGTGTACGCGGTGATGG 613  
QY 1390 GCGCCCGATGGCTCTAGAGTTGCCCTGCTGGTGGCAAGGACAAAGTTATAAGCTATGGGCT 1449  
Db 614 GCACCGGACGGTCCCGCTGCTTCCGGTGGCAAAAGATAAGGTGTTAAACATGTGGGCA 673  
QY 1450 TATTAA 1455  
Db 674 TATTAA 679

RESULT 13  
AL520339  
LOCUS  
DEFINITION AL520339 LTI\_NFL004\_NBC2 Homo sapiens cDNA clone CS0DB006YB09 5  
prime, mRNA sequence.  
ACCESSION AL520339  
VERSION AL520339.1 GI:12783832  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 975)  
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
source  
1. .975  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DB006YB09"  
/clone\_lib="LTI\_NFL004\_NBC2"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco.RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed  
by Life Technologies. Contact : Feng Liang Life  
Technologies, a division of Invitrogen 9800 Medical Center



Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
.8371 Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 195 a 293 c 291 g 184 t 12 others

ORIGIN

Query Match 21.3%; Score 331.6; DB 9; Length 975;  
Best Local Similarity 60.2%; Pred. No. 1.7e-44;  
Matches 532; Conservative 9; Mismatches 343; Indels 0; Gaps 0;

QY 54 GATACAGCGCGCTCGTTTACACGGGCGAGGAAGCCGCCGCCCAATCGACCTGCCGCG 113  
Db 52 GCTAGTCAGTTCAGGATGAGGGCGGCAGCTGCTGGGTCCCGTTCGACGTGCCCGT 111

QY 114 AGGAATCACTACCCAGCAATTGGGACTGATTGCAACGGCTGCTGAAAAACGAGGAAGC 173  
Db 112 GGACATCACCCCGACAGCGTGCAGCTCGTGTGCAACGGCTACTGGCCACGAGGATCC 171

QY 174 CACTCCATATTGTTTTTCGTGGGCGGAGATGAGATCAAGAAGCGCTGGAGGACAGTT 233  
Db 172 CCTGCGCACTGGCTTTCTTKTCCACGATGCTGAGATCGTCTCCTCACTGGGAARACGTT 231

QY 234 GGACTGGCGTCAGTGGACACCGAAACGTCGATCGATATGTTGTATCAGCCACAGCGGT 293  
Db 232 GGAGTCCCAGGCACTGGAGACARARAAGTCTTARACATCATCTACGACCCACAGGCTAT 291

QY 294 TTTCAAAGTGGCCCGCAGTGACAAGATGCAGAGTTTCCATGCCGGGACACGCCGAGGCTGT 353  
Db 292 CTTCARAGTCCGGGCTGTGACTCGCTGCACAGCTCCTTGGAGGTCACAGTGAGCGAGT 351

QY 354 GGTTCGCTGAATTCAGCCCGGATGGTGTCTATCTCGCAGTGGAGTGGACACACCAC 413  
Db 352 CATTTCTGTGGCTTCARCCCTACGGGAARTACCTGGCCAGTGGCTCTGGAGACACCAC 411

QY 414 AGTGGATTTGGGATCTTAACACAGACAGACCGCACTTCACTGTCACAGGTCAATAGCA 473  
Db 412 CGTGGCTTCTGGGATCTCAGACAGACAGACACCACATTTACATGCAAGGGACACARACA 471

QY 474 GTGGGTTCTGTGGGTATCCTGGGCTCCGGATGGCAACCGTTGGCCAGCGTTGCAAGC 533  
Db 472 CTGGGTCCTTAGTATATCCTGGTCTCCAGTGGCAAGAGCTGGCCCTCAGGCTGCAAGAA 531

QY 534 GGGCTCTATAATCATCTGGGACCCCGAGAGGGGTGACAGAAAGGGGCGACCCCTTGAGTG 593  
Db 532 TGGCCAGATTCTCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGGCGAGGACCCCTGCTGG 591

QY 594 GCACAAGAAACACATCAACTGCTCGCTCGGTAACCGTATCATCGCGATCCGGAGTGCAG 653  
Db 592 CCACAGCAAGTGGATCACAGGCTGAGCTGGGAGCCCTCCATCGGAACCCCTGAGTCCG 651

QY 654 GAAACTTGCTTCGGCCAGTGGAGACGCGGACTGCGGATTTGGACCTAATAATTTGGGCCA 713  
Db 652 CTATGTGGCCAGAGCTCCAAGGATGGCAGTGTGCGGATCTGGACACAACATGCAAGCCG 711

QY 714 GTGCCTTATGAACATTTCCGGACACACAATGCTGTGACAGCAGTGAGATGGGTGGAGC 773  
Db 712 CTGTGAGCGCATCTCACCGGGCACACCCAGTCGGTCACTGTCTCCGGTGGGAGGGGA 771

QY 774 GGGCTTATTTATACATCTCCAAAGATCCACAGTGCAGATGTGGGAGCAGCTGATGG 833  
Db 772 CGGGCTTCTCTACTCTGCTCCAGGACCCACCATCAARTCTGGAGAGCTCATGACGG 831

QY 834 AATCTGTGCCGAGCGTTCTCTGGCCAAAGTCACTGGGTAAACAACATTTGCGTGCAGCAC 893  
Db 832 TGTGCTGTGCCGAGACTCTGCAAGGCCACCGCCACTGGGTGAACAACATGGCCCTCARCAC 891

QY 894 CGATTACGTCCTGCGCACTGGTCCATTCCATCCGGTGAAGGATC 937  
Db 892 TGACTATGCCCTGCGCACTGGGGCCCTTTGAACTGCTGAGGCCCTC 935

RESULT 14  
AL526575

LOCUS AL526575 925 bp mRNA linear EST 13-FEB-2001  
DEFINITION AL526575 LTI\_NFL003\_NBC3 Homo sapiens cDNA clone CS0DC014YN14 5  
prime, mRNA sequence.  
ACCESSION AL526575  
VERSION AL526575.1 GI:12790068  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 925)  
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
Location/Qualifiers  
1..925  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DC014YN14"  
/clone\_lib="LTI\_NFL003\_NBC3"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-stranded cDNA was digested with Not I and  
cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
vector. Library was normalized. Library was constructed  
by Life Technologies. Contact : Feng Liang Life  
Technologies, a division of Invitrogen 9800 Medical Center  
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
8371 Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 183 a 278 c 290 g 173 t 1 others

ORIGIN

Query Match 20.2%; Score 314; DB 9; Length 925;  
Best Local Similarity 60.7%; Pred. No. 1.2e-41;  
Matches 529; Conservative 1; Mismatches 341; Indels 1; Gaps 1;

QY 54 GATACAGCGCGCTCGTTTACACGGGCGAGGAAGCCGCCGCCCAATCGACCTGCCGCG 113  
Db 54 GCTAGTCAGTTCAGGATGAGGGCGGCAGCTGCTGGTTCCTCCGTTTCGACGTGCCCGT 113

QY 114 AGGAATCACTACCCAGCAATTGGGACTGATTGCAACCGCTGCTGAAAAACGAGGAAGC 173  
Db 114 GGACATCACCCCGACAGGCTGCAGCTCGTGTGCAACGCGTACTGGCCACGAGGATCC 173

QY 174 CACTCCATATTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTT 233  
Db 174 CCTGCCACTGGCTTTCTTTGTCCACGATGCTGAGATCGTCTCCTCACTGGGGAAGACGTT 233

QY 234 GGACTGGCGTCAGTGGACACCCGAAACGTCGATATTTGTATCAGCCACAGCGGT 293  
Db 234 GGAGTCCCAGGCACTGGAGACAGAGAAGGTCCTAGACATCATCTACCAGCCACAGGCTAT 293

QY 294 TTTCAAAGTGGCCCGCAGTGACACAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGT 353  
Db 294 CTTACAGTCCGGGCTGTGACTCGCTGCACCCAGCTCCTTGGAGGGTCCACAGTGAGGCAGT 353

QY 354 GGTTCGCTGAATTTTCAGCCCGCGAGTGGTGTCTATCTCGCCAGTGGAGTGGCGACACCCAC 413  
Db 354 CATTTCTGTGGCTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGAGACACACCAC 413

QY 414 AGTGCATTTGGGATCTTAACACAGACAGACACGCACTTCACTGCACAGCTCATAAGCA 473  
Db 414 CGTGGCTTCTGGGATCTCAGCAGAGAGACACACATTTTCATATGCAAGGACACAGACA 473

QY 474 GTGGGTTCTGTGCGGATCTCTGGGCTCCGGATGGCAACCGTTGGCCAGCGTTGCAAGC 533



```
Db 474 CTGGGTCTTATATATCTCTGGTCTCCAGATGGCAAGAGCTGGCCTCAGGCTGCAAGAA 533
Qy 534 GGGTCTATATCATCTGGACCCGGAGACGGGTGACAGAGAGGGGACCCCTTGTAGTGG 593
Db 534 TGGCCAGATTCTCTCTGGGACCCCAAGACAGGGAAGGAGGTGGGACGACCCCTCGCTGG 593
Qy 594 GCACAAGAAACACATCAACTGCCTCGCTCGGGAACCGGTATCATCGCATCCGGAGTGCAG 653
Db 594 CCACAGCAAGTGGATCACAGGCTGAGCTGGGAGCCCTCCATGCGGAACCCCTGAGTGCCG 653
Qy 654 GAAACTTGCTTCGCCCAGTGGAGACGGGGACTGCCGGATTGGGACGTAATAATTGGGCCA 713
Db 654 CTATGTGGCCAGCAGCTCCAAAGATGGCAGTGTGCGGATCTGGGACACAACACTGCAGGCCG 713
Qy 714 GTGCTTATGAACATGCGGGACACACAATGCTGTGACAGCAGTGACATGGGTGGAGC 773
Db 714 CTGTAGCGCATCTCTACCGGCACACCCAGTCGGTCACTGTCTCCGGTGGGAGGGGA 773
Qy 774 GGGCTTATTTATACATCTCCAAAGATCCGACAGTGAAGATGTGGGAGCAGCTGATGG 833
Db 774 CGGGTCTCTACTCTGCTCCAGGACCGACCATCAAGTCTGGAGAGCTCATGACGG 833
Qy 834 AATCTGTGCGGAGGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATGCGCTGAGCAC 893
Db 834 TGTGTGTGCGGGAC-TCTGCAAGGCACCGGCCACTGGGTGAACACCATGGCCCTCARGAC 892
Qy 894 CGATTACGTCCTGGGCACCTGGTCCATTCCATC 925
Db 893 TGACTATGCCCTGGGCACCTGGGGCCTTTGAAC 924
```

```
RESULT 15
AL530778
LOCUS AL530778 LTI_NFL001_NBC4 1114 bp mRNA linear EST 13-FEB-2001
DEFINITION AL530778 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD008YA24 5
prime, mRNA sequence.
ACCESSION AL530778
VERSION AL530778.1 GI:12794271
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1114)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..1114
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD008YA24"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/Note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 233 a 320 c 351 g 180 t 30 others
ORIGIN
```

```
Query Match 19.4%; Score 301.4; DB 9; Length 1114;
Best Local Similarity 57.3%; Pred. No. 1.2e-39;
Matches 582; Conservative 19; Mismatches 406; Indels 8; Gaps 4;
Qy 425 GGGATCTTAACACAGACACACCGCACTTCCACCTGCACAGGTTCATAAGCAGTGGTCTGT 484
Db 2 GGGAKCCGACACAGACACACCACTTTACAGGSCRAGGACACAGACACTGGGGCCTTA 61
Qy 485 GCGTATCTCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGGCAAGCGGGCTC-TATA 543
Db 62 GGAGATSCGGGCTCCAGAGGGCAAGARGCGGGCTCAGGCKGCRAGARTGGCGCAGAT 121
Qy 544 ATCATCTGGGACCCGGAGACGGGTGACAGAGAAGGGGCGACCCCTTGAGTGGGCACAGAAA 603
Db 122 CTCCGCTGGGACCCCAAGRCAGGGAAGCAGGKGGCAGGACCCCTCGCKGCCACAGCAAG 181
Qy 604 CACATCAACTGCCTCGCTGGGAACCGTATCATCGCGATCCGGAGTCCGGAACCTTGCT 663
Db 182 TGGATCACAGSSSTGAGCGGGAGCCCTCCATGCGAACCCCTGAGTGCSCGATGTGGCC 241
Qy 664 TCCGCCAGTGGAGACGGGGACTGCCGGATTGGGACGTAAATTTGGGCCAGTGCCTTATG 723
Db 242 AGCAGTCCAAGGATGGCAGTGTGCGGRKSKGGGACACAACTGCAGGCCCTGTGAGCGC 301
Qy 724 AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGCGGSCCTTATT 783
Db 302 ATCKKACCGGGCACACCCAGTGGTCAACCKGTSTCCGGTGGGAGGGGACGGGCTTCTC 361
Qy 784 TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC 843
Db 362 TACHCTGCCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCKCATGACGGTGTGCKGTC 421
Qy 844 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATTTGGCTGAGCACCCGATTACGTC 903
Db 422 CGGACTGTCAAGGCCACGGCCACTGGGTGAACACCATGGCCCTCAGCAGCTACTATGCC 481
Qy 904 CTGGCAGCTGGTCCATTCCATCCGGTGAAGGATCGTCCCAAGGCC--ACCTCAGTTG 960
Db 482 CTGGCAGCTGGGGCCTTTGAACCTGCKGAGGCCCTCACTGAATCCCAAGACCTCCAAGGA 541
Qy 961 AGCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCGGTGTGCCGTGACGAG 1020
Db 542 TCCTTGCAGAGAGTGAAGAGAGAGGGCTCTGAGCCGATACACACCTCGTGGGGCAGGGT 601
Qy 1021 GTGGAGTCGCTGGTTTCTCTGTTGCGATGACAAACACCTCTATCTGTGG--CGGAACAAC 1077
Db 602 CCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACTTATCTCTGGTCCCGCAGAG 661
Qy 1078 CAGAACAAAGTCGTTGAGCGCATGACAGGGCACCAAGCGTGGTCAACGATGTGAATAT 1137
Db 662 GACAAAAAGCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCAGGTCTCTTC 721
Qy 1138 TCGCCGGATGTAAGCTAATTGCGTCTGCTTTCATTGTGACAAAGTCACTGCGTCTGTGGCGA 1197
Db 722 TCTCCTGACTCCCGCATCGTGGCTAGTGCCTCCTTTGACAAAGTCCATCAAGCTGTGGAT 781
Qy 1198 GCCAGCGATGCTCAGTACATGGCCACCTTCCGGGGTCATGTGCAGGCTGTTTACACGGTT 1257
Db 782 GGCAGGACGGGCAAGTACCTGGCTTCCCTACGCGGCCACGTTGGCTGCCGTGTACCAGATT 841
Qy 1258 GCCTGGTCCGGGACTCCCGCTTGTATTGTTTCCGGCAGCAAAAGACTCAACTCTAAAAGTA 1317
Db 842 GCGTGGTCAAGTGCAGTCGGCTCCTGGTCAAGCAGCAGTGCAGCAGCAGTGAAGGTG 901
Qy 1318 TGGAGTGTGACAGCAAGAACTGGCAGAGGAGTGCCTGGACATGCGGATGAGGTGTTT 1377
Db 902 TGGGATGTGAAGGCCCGCCAGAGCTGGCCATGGACCTGCCCGGCCACCGGATGAGGTATAT 961
Qy 1378 GGAGTGGAGTGGCGGCCCGATGGCTCTAGAGTTGCCCTCTGTGGTGGCAAGGACAAG 1432
Db 962 GCTGTTGACTGGAGTCCAGATGGC-CAGAAATRGCAAGTGGTGGGAAGGACAATG 1015
```

Search completed: June 2, 2003, 06:52:31  
Job time : 2088 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 02:56:24 ; Search time 3985 Seconds  
(without alignments)  
11356.302 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattcccaaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1555	100.0	1555	6	AX025192	AX025192 Sequence
2	1528.2	98.3	1663	3	AY089286	AY089286 Drosophil
3	1521	97.8	1521	3	DME012588	AJ012588 Drosophil
4	1285.8	82.7	85095	3	AC004573	AC004573 Drosophil
5	1285.8	82.7	142257	2	AC017491	AC017491 Drosophil
6	1285.8	82.7	182726	3	AC008002	AC008002 Drosophil
7	1285.8	82.7	302473	3	AE003589	AE003589 Drosophil
8	478	30.7	1859	6	AX135811	AX135811 Sequence
9	477.4	30.7	1853	9	AK001320	AK001320 Homo sapi
10	477.4	30.7	1854	6	AX077627	AX077627 Sequence
11	477.4	30.7	1870	9	BC002884	BC002884 Homo sapi
12	477.4	30.7	2593	9	BC012075	BC012075 Homo sapi
13	477.4	30.7	2638	6	AX405715	AX405715 Sequence
14	476.6	30.6	1784	10	BC018399	BC018399 Mus muscu
15	466.2	30.0	1928	6	AX135807	AX135807 Sequence
16	464.6	29.9	2250	5	AF069737	AF069737 Xenopus l
17	249.4	16.0	1170	6	AX135809	AX135809 Sequence
18	210.8	13.6	316613	8	SCCHRII1	X59720 S.cerevisia
19	159.8	10.3	1563	6	AX488920	AX488920 Sequence
20	120.4	7.7	378	6	AX072820	AX072820 Sequence
21	115.8	7.4	40305	8	SPCC18	AL031907 S.pombe c
22	111.6	7.2	1373	9	HSAJ5257	AT005257 Homo sapi
23	90.4	5.8	31787	3	AF106576	AF106576 Caenorhab
24	87.8	5.6	215786	9	AC022916	AC022916 Homo sapi
25	86.2	5.5	203769	2	AL603745	AL603745 Mus muscu
26	86.2	5.5	216655	2	AC120837	AC120837 Mus muscu
27	84.6	5.4	132348	10	CNS07Y02	AL713882 Mus muscu
28	82.8	5.3	334520	1	AP003588	AP003588 Nostoc sp
29	79.8	5.1	34131	2	AC096796	AC096796 Rattus no
30	79.8	5.1	178633	2	AC115671	AC115671 Rattus no
31	79.4	5.1	1844	9	BC026080	BC026080 Homo sapi
32	79.4	5.1	333550	1	AP003593	AP003593 Nostoc sp
33	78	5.0	4722	8	AF323585	AF323585 Podospora
34	77	5.0	4619	1	AF115313	AF115313 Thermomon
35	76.2	4.9	2740	9	HSM804229	AL832918 Homo sapi
36	76.2	4.9	81575	8	AB009055	AB009055 Arabidops
37	74.4	4.8	2311	3	AY113628	AY113628 Drosophil
38	73.6	4.7	5413	8	AF323583	AF323583 Podospora
39	72.8	4.7	1209	3	AF100287	AF100287 Trypanoso
40	72	4.6	5413	8	AF323582	AF323582 Podospora
41	72	4.6	5413	8	AF323584	AF323584 Podospora
42	72	4.6	10280	1	AE012949	AE012949 Chlorobiu
43	71	4.6	600	11	CNS06EEP	AL395015 T3 end of
44	70.6	4.5	5411	8	FLABTLP	L28125 Podospora a
45	68.4	4.4	2977	6	AX430450	AX430450 Sequence

ALIGNMENTS

RESULT 1	AX025192	AX025192	1555 bp	DNA	linear	PAT 16-SEP-2000
LOCUS	Sequence 9 from Patent WO0026364.					
DEFINITION	AX025192					
ACCESSION	AX025192.1	GI:10186913				
VERSION						
KEYWORDS	Drosophila sp.					
SOURCE	Drosophila sp.					
ORGANISM	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.					
REFERENCE	1 (bases 1 to 1555)					
AUTHORS	Cohen,S., Bouwmeester,A. and Royet,J.					
TITLE	Regulator of notch signaling activity					

JOURNAL Patent: WO 0026364-A 9 11-MAY-2000; EUROP MOLECULAR BIOLOGY LAB (DE) ; COHEN STEPHEN (DE) ; BOUWMEESTER ANTONIUS (DE) ; ROYET JULIEN (FR) FEATURES Location/Qualifiers source 1. .1555 /organism="Drosophila sp." /db_xref="taxon:7242" BASE COUNT 418 a 373 c 448 g 316 t ORIGIN									
Query Match 100.0%; Score 1555; DB 6; Length 1555; Best Local Similarity 100.0%; Pred. No. 0; Matches 1555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	AATTCCCAAAAAATGCAGGAGACGACACGAGCAAGGCGCCACGCCACATACGATACAG	60						
Db	1	AATTCCCAAAAAATGCAGGAGACGACACGAGCAAGGCGCCACGCCACATACGATACAG	60						
QY	61	GCGCGCCTCGTTTACACGGGCGAGGAAGCCGCCCAATCGACCTGCCGCGAGGAATC	120						
Db	61	GCGCGCCTCGTTTACACGGGCGAGGAAGCCGCCCAATCGACCTGCCGCGAGGAATC	120						
QY	121	ACTACCCAGCAATGGGACTGATTTGCAACGGCGTCTGTAACGAGGAGGAGCCACTCCA	180						
Db	121	ACTACCCAGCAATGGGACTGATTTGCAACGGCGTCTGTAACGAGGAGGAGCCACTCCA	180						
QY	181	TATTTGTTTTTTCGTGGGCGAGGATGAGATCGATATTGTGTATCAGCCACAGCGCGTTT	240						
Db	181	TATTTGTTTTTTCGTGGGCGAGGATGAGATCGATATTGTGTATCAGCCACAGCGCGTTT	240						
QY	241	CGCTCAGTGGACACCCGAAAACGTCGATCGATATTGTGTATCAGCCACAGCGCGTTT	300						
Db	241	CGCTCAGTGGACACCCGAAAACGTCGATCGATATTGTGTATCAGCCACAGCGCGTTT	300						
QY	301	GTGGCGCCAGTGACAAAGATGCACGAGTTCATCGCCAGTGGAGTGGCGAGCGTGTGG	360						
Db	301	GTGGCGCCAGTGACAAAGATGCACGAGTTCATCGCCAGTGGAGTGGCGAGCGTGTGG	360						
QY	361	CTGAATTTACGCCCGGATGCTGCTCATCTCGCCAGTGGAGTGGAGTGGCGAGTGGG	420						
Db	361	CTGAATTTACGCCCGGATGCTGCTCATCTCGCCAGTGGAGTGGAGTGGCGAGTGGG	420						
QY	421	TTGTGGGATCTTAACACAGAGACACCCGCACTTCACCTGCACAGGTCATAAGCAGTGG	480						
Db	421	TTGTGGGATCTTAACACAGAGACACCCGCACTTCACCTGCACAGGTCATAAGCAGTGG	480						
QY	481	CTGTGCGTATCCTGGGCTCGGATGGCAACCGTTTGGCCAGCGGTTGCAAGCGGCTCT	540						
Db	481	CTGTGCGTATCCTGGGCTCGGATGGCAACCGTTTGGCCAGCGGTTGCAAGCGGCTCT	540						
QY	541	ATAATCATCTGGGACCCGGAGACGGGTCAGCAGAAAGGCGCGACCCCTTGAGTGGG	600						
Db	541	ATAATCATCTGGGACCCGGAGACGGGTCAGCAGAAAGGCGCGACCCCTTGAGTGGG	600						
QY	601	AAACACATCAACTGCCCTCGCCTGGGAACCGTATCATCCGATCCGGAGTGCAGGAACTT	660						
Db	601	AAACACATCAACTGCCCTCGCCTGGGAACCGTATCATCCGATCCGGAGTGCAGGAACTT	660						
QY	661	GCTTCCGCCAGTGGAGACGGGACTGCCGGATTGGGACGTAAATTTGGCCAGTGCCTT	720						
Db	661	GCTTCCGCCAGTGGAGACGGGACTGCCGGATTGGGACGTAAATTTGGCCAGTGCCTT	720						
QY	721	ATGAACATTCGGGACACACAAATGCTGTGACAGCAGTGAATGGGTTGGAGCGGCTT	780						
Db	721	ATGAACATTCGGGACACACAAATGCTGTGACAGCAGTGAATGGGTTGGAGCGGCTT	780						
QY	781	ATTTATACATCCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGAATCTTG	840						
Db	781	ATTTATACATCCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGAATCTTG	840						
QY	841	TGCCGGACGTTCTCTGGCCCAAGCTCACTGGGTAACAACATTTGCGCTGAGCACCGG	900						
Db	841	TGCCGGACGTTCTCTGGCCCAAGCTCACTGGGTAACAACATTTGCGCTGAGCACCGG	900						

QY	901	GTCTTGGCCTGCTCCATTCCTCCATCCCGTGAAGGATCGCTCCAAGAGCCACCTCAGTTG	960						
Db	901	GTCTTGGCCTGCTCCATTCCTCCATCCCGTGAAGGATCGCTCCAAGAGCCACCTCAGTTG	960						
QY	961	AGCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGGTACCAAGCCGCTGTGCCCTGACGAG	1020						
Db	961	AGCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGGTACCAAGCCGCTGTGCCCTGACGAG	1020						
QY	1021	GTGGAGTCGCTGGTTTCTCTGTTCCGATGACAACACCCCTCTATCTGTGGCGGAACAACAG	1080						
Db	1021	GTGGAGTCGCTGGTTTCTCTGTTCCGATGACAACACCCCTCTATCTGTGGCGGAACAACAG	1080						
QY	1081	AACAAGTCGCTTGAGCGCATGACAGGGCACCAGAACGTCGGTCAACCATGTGAATATTCG	1140						
Db	1081	AACAAGTCGCTTGAGCGCATGACAGGGCACCAGAACGTCGGTCAACCATGTGAATATTCG	1140						
QY	1141	CCGGATGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTGCCTGTGGCGAGCC	1200						
Db	1141	CCGGATGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTGCCTGTGGCGAGCC	1200						
QY	1201	AGCGATGGTCAGTACATGGCCACCTCCGGGTCATGTGCAGGCTGTTTACACGGTTGCC	1260						
Db	1201	AGCGATGGTCAGTACATGGCCACCTCCGGGTCATGTGCAGGCTGTTTACACGGTTGCC	1260						
QY	1261	TGTCGCGGACTCCCGCTTGATTTCCGGCAGCAAGACTCAACTCTAAAAGTATGG	1320						
Db	1261	TGTCGCGGACTCCCGCTTGATTTCCGGCAGCAAGACTCAACTCTAAAAGTATGG	1320						
QY	1321	AGTGTGCAGACGAAAGAACTGGCACAGGAGTGCCTGGCAAGGACAAAGTATATAAG	1380						
Db	1321	AGTGTGCAGACGAAAGAACTGGCACAGGAGTGCCTGGCAAGGACAAAGTATATAAG	1380						
QY	1381	GTGACTGGCGCCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGACAAAGTATATAAG	1440						
Db	1381	GTGACTGGCGCCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGACAAAGTATATAAG	1440						
QY	1441	CTATGGGCTTATTAACAAATCATTAACTTGACACGGTAAGAAAATACTTAGGAATAAAG	1500						
Db	1441	CTATGGGCTTATTAACAAATCATTAACTTGACACGGTAAGAAAATACTTAGGAATAAAG	1500						
QY	1501	TAAACGTCCTGAGTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1555						
Db	1501	TAAACGTCCTGAGTAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1555						

RESULT 2  
AY089286

LOCUS AY089286 1663 bp mRNA linear INV 18-MAR-2002  
DEFINITION Drosophila melanogaster AT08344 full insert cdna.

ACCESSION AY089286

VERSION AY089286.1 GI:19527818

KEYWORDS FLI\_CDNA.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1663)

AUTHORS

Stapleton,M., Brokstein,P., Hong,L., Aghayani,A., Carlson,J.,  
Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,  
George,R., Gonzales,M., Guarin,H., Kronmiller,B., Li,P., Liao,G.,  
Miranda,A., Mungall,C.J., Nunoo,J., Pacleeb,J., Paragas,V., Park,S.,  
Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.  
and Celniker,S.

TITLE Direct Submission

JOURNAL Submitted (12-MAR-2002) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA

COMMENT Sequence submitted by:

Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720



This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).

FEATURES		Location/Qualifiers	
source		1..1663	/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"	
gene		/map="21C6-21C6"	
		1..1663	
CDS		/gene="Nle"	
		/note="alignment with genomic scaffold AE003589"	
		/db_xref="FLYBASE:FBgn0021874"	
		100..1566	
		/gene="Nle"	
		/note="Longest ORF"	
		/codon_start=1	
		/product="At08344p"	
		/protein_id="AAL90024.1"	
		/db_xref="GI:19527819"	
BASE COUNT	444 a	397 c	478 g
		344 t	
ORIGIN		/translation="MLAKKQMQETDTEQEA TPHTIQARLVSDTGEEAGPPIDLPA GI TTQQLGLICNALNKNEATPYLFFVGEDEIKKSLEDTLDLASVDTENVIDIVYQPAV FKVPRVTRCTSSMPGHAENVSLNFS PDGAHLASGSDTTVRLWDLNTEPHFTCTGH KQWVLCVSWAPDGKRLASGCKAGSII IWDPETQOQGRPLSGHKHKHINCLAWEPYHRD PECKRLASAGDGCRIWDVKLQCLMNIAGHTNAVTVRWGGAGLIYTSKDRTVKM WRAADGILCRTFSGHAHWVNIALSTDYVLRTPFPHPVKDRSKSHLSLSTELQESAL KRYQAVCPDEVESLVSCDDNTLYLWRNNQKVERMTGHQNVVNDVKYSPDVKLLIAS ASFQSVRLWRASDGQYMATFRGHVQAVTYTVASADSRLIVSGSKDSTLLKVMVSQTKK LAQELPGHADEVFGVDWAPDGSRVASGGKDKVIKLMAY"	
		Query Match 98.3%; Score 1528.2; DB 3; Length 1663;	
		Best Local Similarity 99.6%; Pred. No. 0;	
		Matches 1543; Conservative 0; Mismatches 3; Indels 3; Gaps 1;	
		QY	7 CAAAAATGCAGGACGGACAGGAGCAAGAGGCCACGCCACATACGATACAGCGCGC 66
		Db	115 CAGAAATGCAGGACGGACAGGAGCAAGAGGCCACGCCACATACGATACAGCGCGC 174
		QY	67 CTCGTTT---ACACGGCGGAGGAAGCGCGCCGCCCAATCGACCTGCCGGCAGGAATCACT 123
		Db	175 CTCGTTTCGGACACGGCGGAGGAAGCGCGCCGCCCAATCGACCTGCCGGCAGGAATCACT 234
		QY	124 ACCCAGCAATTGGGACTGATTTGCAACCGCGTGTGTAACCAAGCAAGCAAGCCACTCCATAT 183
		Db	235 ACCCAGCAATTGGGACTGATTTGCAACCGCGTGTGTAACCAAGCAAGCAAGCCACTCCATAT 294
QY		184 TTGTTTTTCGTGGCGAGGATGAGATCAAGAGACGCTGGAGACACGTTGGACTTGGCG 243	
		Db	295 TTGTTTTTCGTGGCGAGGATGAGATCAAGAGACGCTGGAGACACGTTGGACTTGGCG 354
QY		244 TCAGTGGACACCGAAACCGTGATCGATATTGTGTATCAGCCACAGCGCGTTTCAAAGTG 303	
Db		355 TCAGTGGACACCGAAACCGTGATCGATATTGTGTATCAGCCACAGCGCGTTTCAAAGTG 414	
QY		304 CGCCCACTGACAAGATGCACGAGTCCATGCCGGACACGCCGAGGCTGTGGTTTCGCTG 363	
Db		415 CGCCCACTGACAAGATGCACGAGTCCATGCCGGACACGCCGAGGCTGTGGTTTCGCTG 474	
QY		364 AATTTACGCCCGGATGGTGTCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTCGGATTG 423	

Db	475 AATTTACGCCCGGATGGTGTCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTCGGATTG 534
QY	424 TGGGATCTTAACACAGACAGACACCGCACCTTCACCTGCACAGGTGTAAGCAGTGGGTTCTG 483
Db	535 TGGGATCTTAACACAGACAGACACCGCACCTTCACCTGCACAGGTGTAAGCAGTGGGTTCTG 594
QY	484 TCGGTATCCTGGGCTCCGGATGGCAAAACGGTGGCCAGCGGTTGCAAGCGGGCTCTATA 543
Db	595 TCGGTATCCTGGGCTCCGGATGGCAAAACGGTGGCCAGCGGTTGCAAGCGGGCTCTATA 654
QY	544 ATCATCTGGGACCCGGAGACGGGTGAGCAGAGAGGGGCGACCCCTTGAGTGGGCACAAAGAAA 603
Db	655 ATCATCTGGGACCCGGAGACGGGTGAGCAGAGAGGGGCGACCATTGAGTGGGCACAAAGAAA 714
QY	604 CACATCAACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGCT 663
Db	715 CACATCAACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGCT 774
QY	664 TCCGCCAGTGGAGACGGGACTGCCGGATTGGGACGTAAATTTGGGCCAGTCGCCTTATG 723
Db	775 TCCGCCAGTGGAGACGGGACTGCCGGATTGGGACGTAAATTTGGGCCAGTCGCCTTATG 834
QY	724 AACATTTCCCGGACACACAAATGCTGTGACAGAGTGAGATGGGGTGGAGCGGCGCCTTATT 783
Db	835 AACATTTCCCGGACACACAAATGCTGTGACAGAGTGAGATGGGGTGGAGCGGCGCCTTATT 894
QY	784 TATACATCCTCCAAAGATCGCAGAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC 843
Db	895 TATACATCCTCCAAAGATCGCAGAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC 954
QY	844 CGGAGCTTCTCTGCCCAAGCTCACTGGGTTAAACAACATTCGCGTGAGCACCCGATTACGTC 903
Db	955 CGGAGCTTCTCTGCCCAAGCTCACTGGGTTAAACAACATTCGCGTGAGCACCCGATTACGTC 1014
QY	904 CTGCGCACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC 963
Db	1015 CTGCGCACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC 1074
QY	964 ACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTCCTTGACGAGGTG 1023
Db	1075 ACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTCCTTGACGAGGTG 1134
QY	1024 GAGTCGCTGGTTTCCTGTTCCGATGACAACACCCCTCTATCTGTGGCGGAACAACAGAAC 1083
Db	1135 GAGTCGCTGGTTTCCTGTTCCGATGACAACACCCCTCTATCTGTGGCGGAACAACAGAAC 1194
QY	1084 AAGTGCCTTGAGCGCATGACAGGGCACAGAACGTTGTTCAACGATGTGAAATATTCGCCG 1143
Db	1195 AAGTGCCTTGAGCGCATGACAGGGCACAGAACGTTGTTCAACGATGTGAAATATTCGCCG 1254
QY	1144 GATGTAAGCTAATTGCGTCTGCTTCATTTGACAATFCAGTCGTCGTGTGGCGAGCCAGC 1203
Db	1255 GATGTAAGCTAATTGCGTCTGCTTCATTTGACAATFCAGTCGTCGTGTGGCGAGCCAGC 1314
QY	1204 GATGTCAGTACATGGCCACCTTCCGGGGTTCATGTGAGGCTGTTTACACGGTTGCCTGG 1263
Db	1315 GATGTCAGTACATGGCCACCTTCCGGGGTTCATGTGAGGCTGTTTACACGGTTGCCTGG 1374
QY	1264 TCCGGGACTCCCGCTTGAATGTTTCCGGCAGCAAAAGACTCAACTCTAAAAGTATGGAGT 1323
Db	1375 TCCGGGACTCCCGCTTGAATGTTTCCGGCAGCAAAAGACTCAACTCTAAAAGTATGGAGT 1434
QY	1324 GTGCAGACGAAGAAACTGGCACAGGAGTGCCTGGACATCGGGATGAGGTGTTCCGGAGTG 1383
Db	1435 GTGCAGACGAAGAAACTGGCACAGGAGTGCCTGGACATCGGGATGAGGTGTTCCGGAGTG 1494
QY	1384 GACTGGGCGCCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGACAAAGTTATAAAGCTA 1443
Db	1495 GACTGGGCGCCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGACAAAGTTATAAAGCTA 1554
QY	1444 TGGGCTTATTAAACAAATCAATTAACCTTGACACGGTAAGAAAATCTTAGGAATAAAGTAA 1503

Db 1555 TGGGCTTATTACAAATCATTAACCTGTACACGGTAAGAAATACTTAGGAATAAGTAA 1614

QY 1504 AACGTCCTGAGTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1552

Db 1615 AACGTCCTGAGTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1663

RESULT 3

DME012588

LOCUS DME012588 1521 bp mRNA linear INV 12-MAR-1999

DEFINITION Drosophila melanogaster mRNA for Notchless protein.

ACCESSION AJ012588

VERSION AJ012588.1 GI:4127780

KEYWORDS notchless gene; WD40 repeats.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1521)

AUTHORS Royet,J.

TITLE Direct Submission

JOURNAL Submitted (28-OCT-1998) Royet J., Developmental Biology Programme, EMBL, Meyerhofstrasse 1 Heidelberg, 69117, GERMANY

REFERENCE 2 (bases 1 to 1521)

AUTHORS Royet,J., Bouwmeester,T. and Cohen,S.M.

TITLE Notchless encodes a novel WD40-repeat-containing protein that modulates Notch signaling activity

JOURNAL EMBO J. 17 (24), 7351-7360 (1998)

MEDLINE 99077802

PUBMED 98571191

FEATURES

source Location/Qualifiers

1..1521

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

1..1443

/gene="notchless"

1..1443

/gene="notchless"

/note="WD40 repeat containing protein"

/codon\_start=1

/product="Notchless protein"

/protein\_id="CAA10070.1"

/db\_xref="GI:4127781"

/translation="MQETDTEQEATPTIQARLVYTGEEAGPPIDLPAGITQQQLGLI CNALKNEEATPYLFFVGEDEIKKSLEDLTLASVDTEINVIDIVYQPAVFKVRPVTR CTSSMPGHAENVSLNFSPDGAHLASGSDTVRLWDLNLTETPHFTCTGKHQWVLCVS WAPDKRLASGKAGSIIIWDETQQKGRPLSGHKHKLINCLAWEPYHRDPECKRLAS ASGDGCRIMVDVLGQCLMNIAGHTNAVTVRWGGAGLLIYTSKDRYVKMWRAADGIL CRTFSGQAHVNNIALSTDYLRTPFPHPVKDRSKSHLSLSTEELOESALKRYQAVCP DEVESLVSCDDNTLYLRNKNKCVERTMGHQNVDVYKSPDVKLIASASFDKSVR LWRASDGOYMAFERGHVQAVTVNWSADSRLLVSGSKDSLKWSVQTKKLAQELPGH ADEVFGVDWAPDGSRVASGGKDKVIKLWAY"

BASE COUNT 389 a 370 c 448 g 314 t

ORIGIN

Query Match 97.8%; Score 1521; DB 3; Length 1521;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ATGCAGGAGACGGACACGGAGCAAGAGCCACGCACATACGATACAGGCGCGCTCGTT 72

Db 1 ATGCAGGAGACGGACACGGAGCAAGAGCCACGCACATACGATACAGGCGCGCTCGTT 60

QY 73 TACACGGCGAGGAAGCCCGCCGCAATCGACCTGCCGGCAGGAATCACTACCCAGCAA 132

Db 61 TACACGGCGAGGAAGCCCGCCGCAATCGACCTGCCGGCAGGAATCACTACCCAGCAA 120

QY 133 TTGGGACTGATTGCAACGGCTGCTGAAAAACGAGGAGCCACTCCATATTTGTTTTC 192

Db 121 TTGGGACTGATTGCAACGGCTGCTGAAAAACGAGGAGCCACTCCATATTTGTTTTC 180

QY 193 GTGGCGGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGCTTGGACTTGGCGTCAGTGGAC 252

Db 181 GTGGCGAGGATGAGATCAAGAAGAGCCTGGAGACACGTTGGACTTGGCGTCAGTGGAC 240

QY 253 ACCGAAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTGGCCCCAGTG 312

Db 241 ACCGAAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTGGCCCCAGTG 300

QY 313 ACAAGATGCACGAGTTCCTATGCCGGGACACGCCGAGGCTGTGGTTTCGCTGAATTCAGC 372

Db 301 ACAAGATGCACGAGTTCCTATGCCGGGACACGCCGAGGCTGTGGTTTCGCTGAATTCAGC 360

QY 373 CCGGATGGTGTCTCATCTGCCAGTGAAGTGGCAGACACACAGTCCGATTGTGGGATCTT 432

Db 361 CCGGATGGTGTCTCATCTGCCAGTGAAGTGGCAGACACACAGTCCGATTGTGGGATCTT 420

QY 433 AACACAGACACCCGACTTCACCTGCACAGGTGATAGCAGTGGGTTCTGTGCGTATCC 492

Db 421 AACACAGACACCCGACTTCACCTGCACAGGTGATAGCAGTGGGTTCTGTGCGTATCC 480

QY 493 TGGGCTCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAGCGGGCTCTATATCATCTGG 552

Db 481 TGGGCTCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAGCGGGCTCTATATCATCTGG 540

QY 553 GACCCGAGACGGGTGACAGAGAAGGGCGACCCCTGAGTGGGACAAAGAACACATCAAC 612

Db 541 GACCCGAGACGGGTGACAGAGAAGGGCGACCCCTGAGTGGGACAAAGAACACATCAAC 600

QY 613 TGCCTCGCCTGGGAACCGTATCATCGCGATCCGAGTGCAGGAACCTTGTCTCCGCCAGT 672

Db 601 TGCCTCGCCTGGGAACCGTATCATCGCGATCCGAGTGCAGGAACCTTGTCTCCGCCAGT 660

QY 673 GGAGACGGGACTGCCGGATTGGGACGTAATAATGGGCCAGTGCCTTATGAACATTGCC 732

Db 661 GGAGACGGGACTGCCGGATTGGGACGTAATAATGGGCCAGTGCCTTATGAACATTGCC 720

QY 733 GGACACACAAATGCTGTGACAGCAGTGAATGGGGTGGAGCGGCGCTTATTTATACATCC 792

Db 721 GGACACACAAATGCTGTGACAGCAGTGAATGGGGTGGAGCGGCGCTTATTTATACATCC 780

QY 793 TCCAAAGATGCACAGTGAAGATGTGGCGAGCAGTGAATGGGAACTTGTGCGGACGTTTC 852

Db 781 TCCAAAGATGCACAGTGAAGATGTGGCGAGCAGTGAATGGGAACTTGTGCGGACGTTTC 840

QY 853 TCTGGCAAAGCTCACTGGGTAAACAACATTGCGGTGAGCACCCGATACGTCCTGCGCACT 912

Db 841 TCTGGCAAAGCTCACTGGGTAAACAACATTGCGGTGAGCACCCGATACGTCCTGCGCACT 900

QY 913 GGTCCATCCATCCGTGAAGATCGCTCCAAGAGCCACCTCAGTTTGAGCACTGAGGAA 972

Db 901 GGTCCATCCATCCGTGAAGATCGCTCCAAGAGCCACCTCAGTTTGAGCACTGAGGAA 960

QY 973 TTGCAGGAATCTGCCTGAAGCGCTACACAGCGCTGTCCTGACGAGGTGAGTGCCTG 1032

Db 961 TTGCAGGAATCTGCCTGAAGCGCTACACAGCGCTGTCCTGACGAGGTGAGTGCCTG 1020

QY 1033 GTTTCCTGTTCCGATGACACACCCCTCTATCTGTGGCGGAACACCAACAGTGCCTT 1092

Db 1021 GTTTCCTGTTCCGATGACACACCCCTCTATCTGTGGCGGAACACCAACAGTGCCTT 1080

QY 1093 GAGCGCATGACAGGGCACCAGAACCTGGTCAACGATGTGAATATTCGCCGATGTAAAG 1152

Db 1081 GAGCGCATGACAGGGCACCAGAACCTGGTCAACGATGTGAATATTCGCCGATGTAAAG 1140

QY 1153 CTAATTGCGTCTGCTTCATTTGACAAGTCAAGTGGTCTGTGGGAGCCAGCATGGTTCAG 1212

Db 1141 CTAATTGCGTCTGCTTCATTTGACAAGTCAAGTGGTCTGTGGGAGCCAGCATGGTTCAG 1200

QY 1213 TACATGGCACCTTCCGGGGTCAATGTCAGGCTGTTTACACGGTTCCTGTCGCGGAC 1272

Db 1201 TACATGGCACCTTCCGGGGTCAATGTCAGGCTGTTTACACGGTTCCTGTCGCGGAC 1260

QY 1273 TCCCGCTTGATTGTTCCGGCAGCAAGAACTCAACTCTAAAAGTATGGAGTGCAGACG 1332

Db 1261 TCCCGCTGATTGTTTCCGGCAGCAAACTCAACTCTAAAGTAGTGAGTGTCAGACG 1320

QY 1333 AAGAACTGGCACAGGAGTGCCTGGACATCGGATGAGGTGTTCCGAGTGGACTGGCG 1392

Db 1321 AAGAACTGGCACAGGAGTGCCTGGACATCGGATGAGGTGTTCCGAGTGGACTGGCG 1380

QY 1393 CCCGATGGCTCTAGAGTTGCCTCTGTGGTCAAGGACAAAGTTATATAAGCTATGGGCTTAT 1452

Db 1381 CCCGATGGCTCTAGAGTTGCCTCTGTGGTCAAGGACAAAGTTATATAAGCTATGGGCTTAT 1440

QY 1453 TAACAAATCATTAACCTTGTCACACGGTAAGAAAAATACTTAGGAATAAAACGTCCTG 1512

Db 1441 TAACAAATCATTAACCTTGTCACACGGTAAGAAAAATACTTAGGAATAAAACGTCCTG 1500

QY 1513 AGTAAAAAANAANAANA 1533

Db 1501 AGTAAAAAANAANAANA 1521

RESULT 4

AC004573/c

LOCUS

DEFINITION

Drosophila melanogaster, chromosome 2L, region 21C5-21D1, Pl clone

DS07610, complete sequence.

ACCESSION

AC004573

VERSION

AC004573.1

KEYWORDS

GI:4204255

SOURCE

HTG.

ORGANISM

Drosophila melanogaster.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 85095)

Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,

Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,

Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,

Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,

Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and

Rubin,G.M.

Sequencing of Drosophila chromosome 2L, region 21C5-21D1

Unpublished (1998)

2 (bases 1 to 85095)

Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,

Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,

Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,

Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,

Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,

Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch,E.,

Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,

Zieran,L.L. and Kimmel,B.E.

Direct Submission

Submitted (17-APR-1998) Drosophila Genome Center, Lawrence Berkeley

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Jan 30, 1999 this sequence version replaced gi:4164118.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send email

to [bdgpe@fruitfly.berkeley.edu](mailto:bdgpe@fruitfly.berkeley.edu).

Pl library location: 80-26.

Location/Qualifiers

1. 85095

/organism="Drosophila melanogaster"

/strain="y2; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="2L"

/map="21C5-21D1"

FEATURES

source

/clone="p1 DS07610 (D275)"

/clone\_lib="p1 library, partial Sau3A in pNS582tet14ad10"

BASE COUNT 22289 a 19671 c 20319 g 22816 t

ORIGIN

Query Match 82.7%; Score 1285.8; DB 3; Length 85095;

Best local Similarity 89.2%; Pred. No. 5.5e-311;

Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;

QY 7 CAAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGAGCGCGC 66

Db 27639 CAGAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGAGCGCGC 27580

QY 67 CTGCTTT---ACACGGGCGAGGAAGCCGGCCGCCCAATCGACTGCCGGCAGGAATCACT 123

Db 27579 CTCGTTTCGGACACGGGCGAGGAAGCGGGCCGCCCAATCGACTGCCGGCAGGAATCACT 27520

QY 124 ACCCAGCAATTTGGACTGATTTGCAACCGCGTCTGCTGAAAAACGAGGAAGCCACTCCATAT 183

Db 27519 ACCCAGCAATTTGGCTCTGATTTGCAACCGCGTCTGCTGAAAAACGAGGAAGCCACTCCATAT 27460

QY 184 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAAGCCCTGGAGGACACAGTTGGACTTGGCG 243

Db 27459 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAAGCCCTGGAGGACACAGTTGGACTTGGCG 27400

QY 244 TCAGTGGACACCGAAAAACGTGATCGATATTTGTATCAGCCACACAGCGGCTTTTCAAAGTG 303

Db 27399 TCAGTGGACACCGAAAAACGTGATCGATATTTGTATCAGCCACACAGCGGCTTTTCAAAGTG 27340

QY 304 CGCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 363

Db 27339 CGCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 27280

QY 364 AATTTACGCCCCGGATGGTGTCTCATCTGCCAGTGGGAAGTGGCGACACACAGTGGGATTG 423

Db 27279 AATTTACGCCCCGGATGGTGTCTCATCTGCCAGTGGGAAGTGGCGACACACAGTGGGATTG 27220

QY 424 TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCTAAGCAGTGGGTTCTG 483

Db 27219 TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCTAAGCAGTGGGTTCTG 27160

QY 484 TCGGTATCTCTGGGCTCCGGATGGCAACCGGTTGGCCAGCGGTTGCAAGCGGGCTCTATA 543

Db 27159 TCGGTATCTCTGGGCTCCGGATGGCAACCGGTTGGCCAGCGGTTGCAAGCGGGCTCTATA 27100

QY 544 ATCATCTGGGACCCGGAGACGGGTTCAGCAGAAGGGCGACCCCTTGAGTGGGCACACAGAAA 603

Db 27099 ATCATCTGGGACCCGGAGACGGGTTCAGCAGAAGGGCGACCCCTTGAGTGGGCACACAGAAA 27040

QY 604 CACATCAACTGCCTCGCTCGCTGGGAACCGGTATCATCGCATCCGGAGTCCGAGGAACCTGCT 663

Db 27039 CACATCAACTGCCTCGCTCGCTGGGAACCGGTATCATCGCATCCGGAGTCCGAGGAACCTGCT 26980

QY 664 TCCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAAATAATGGGCCAGTGCCTTATG 723

Db 26979 TCCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAAATAATGGGCCAGTGCCTTATG 26920

QY 724 AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGGTGGAGCGGCTTAT 783

Db 26919 AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGGTGGAGCGGCTTAT 26860

QY 784 TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGACAGCTGATGGAATCTTGTGC 843

Db 26859 TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGACAGCTGATGGAATCTTGTGC 26800

QY 844 CGGACGTTCTCTGGCCCAAGCTCACTGGGTAAACAACATTCGCTGAGCACCAGATTACGTC 903

Db 26799 CGGACGTTCTCTGGCCCAAGCTCACTGGGTAAACAACATTCGCTGAGCACCAGTACGTC 26740

QY 904 CTGGCGCACTGGTCCATTCCATCCGGTGAAGATCGCTCCCAAGAGCCACCTCAGTTTGTAGC 963

Db 26739 CTGGCGCACTGGTCCATTCCATCCGGTGAAGATCGCTCCCAAGAGCCACCTCAGTTTGTAGC 26680



QY 964 A-----C 965  
|  
Db 26579 AGTAAGAATTCATATGCTCTCTATATATGCTGTATATCCTAAAGAGATTGTTTACAGC 26620  
|  
QY 966 TGAGGAATTGCAGGAATCTGCTTGAAGCGGTACCAGGCGGTGTCCTTGACGAGGTGGA 1025  
|  
Db 26519 TGAGGAATTGCAGGAATCTGCTTGAAGCGGTACCAGGCGGTGTCCTTGACGAGGTGGA 26560  
|  
QY 1026 GTCGCTGGTTTCCTGTTTCGGATGACAACACCTCTATCTGTGGCGGAACACCAAGAACAA 1085  
|  
Db 26559 GTCGCTGGTTTCCTGTTTCGGATGACAACACCTCTATCTGTGGCGGAACACCAAGAACAA 26500  
|  
QY 1086 GTGCGTTGAGCGCATGACAGGACACGAGCGTGGTCAACGATGTGAATATTCGCCGGA 1145  
|  
Db 26499 GTGCGTTGAGCGCATGACAGGACACGAGCGTGGTCAACGACGTGAATATTCGCCGGA 26440  
|  
QY 1146 TGTAAGCTAATTGCGTCTGCTTCATTTGACAAAGTCAGTGGCTGTGTGGCGAGCCAGCGA 1205  
|  
Db 26439 TGTAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTGGCTGTGTGGCGAGCCAGCGA 26380  
|  
QY 1206 TGGTCAGTACATGGCCACCTTCGGGGTCAATGTGCAGGCTGTTTACACGGTGGCTGGTC 1265  
|  
Db 26379 TGGTCAGTACATGGCCACCTTCGGGGTCAATGTGCAGGCTGTTTACACGGTGGCTGGTC 26320  
|  
QY 1266 CGCGGACTCCCGCTTGATTGTTTCGGCGAGCAAGACTCAACTCTAAA----- 1313  
|  
Db 26319 CGCGGACTCCCGCTTGATTGTTTCGGCGAGCAAGACTCAACTCTAAAAGGTTAGCACGT 26260  
|  
QY 1314 -----AGTATGGAGTGTGCAGA 1330  
|  
Db 26259 AGTTTCAGTTTCTAGCGCTAGCTAATACATATCCTCTTGCTTAGTATGGAGTGTGCAGA 26200  
|  
QY 1331 CGAAGAACTGGCACAGGAGCTGCCTGGACATCGCGATGAGGTGTTTCGGAGTGGACTGGG 1390  
|  
Db 26199 CGAAGAACTGGCACAGGAGCTGCCTGGACATCGCGATGAGGTGTTTCGGAGTGGACTGGG 26140  
|  
QY 1391 CGCCCGATGGCTCTAGAGTTGGCTCTGGTGGAAGGACAAAGTTATAAGCT----- 1442  
|  
Db 26139 CGCCCGATGGCTCTAGAGTTGGCTCTGGTGGAAGGACAAAGTTATAAGCTGTAAGTTT 26080  
|  
QY 1443 -----ATGGGCTTA 1451  
|  
Db 26079 GAGTAACGAGGTTTCTATAACAAATATACAGACTCTAATTTGTATCTTTTCAGATGGGCTTA 26020  
|  
QY 1452 TTAACAAATCATTAACCTTGACACGGTGAAGAAATACTTAGGAATAAAGTAAACGTCCT 1511  
|  
Db 26019 TTAACAAATCATTAACCTTGACACGGTGAAGAAATACTTAGGAATAAAGTAAACGTCCT 25960  
|  
QY 1512 GAGTAAA 1518  
|  
Db 25959 GAGTAAA 25953  
|

RESULT 5  
AC017491  
LOCUS  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AC017491  
VERSION AC017491.1 Gi:6553495  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 142257)  
ADAMS, M. and Venter, J.C.  
Direct Submission  
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
This sequence was identified as CDM:10210988 by the submitter.  
For more information on this record e-mail to fly@celera.com.

\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
source  
1. 142257  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 38086 a 33712 c 32464 g 37995 t  
ORIGIN  
Query Match 82.7%; Score 1285.8; DB 2; Length 142257;  
Best Local Similarity 89.2%; Pred. No. 5.9e-311;  
Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;  
QY 7 CAAAAATGCAGGAGACGGACACACGGAGCAAGAGGCCACGCCACATACGATACAGCGCGC 66  
|  
Db 100034 CAGAAATGCAGGAGACGGACACACGGAGCAAGAGGCCACGCCACATACGATACAGCGCGC 100093  
|  
QY 67 CTCGTTT---ACACGGCGAGGAAGCGCGCCGCCCAATCGACCTGCCGGCAGGAATCACT 123  
|  
Db 100094 CTCGTTTCGGACACGGCGAGGAAGCGCGCCGCCCAATCGACCTGCCGGCAGGAATCACT 100153  
|  
QY 124 ACCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACAGGAAGCCACTCCATAT 183  
|  
Db 100154 ACCAGCAATTGGGCTGATTTGCAACGCGCTGCTGAAAAACAGGAAGCCACTCCATAT 100213  
|  
QY 184 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 243  
|  
Db 100214 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 100273  
|  
QY 244 TCAGTGGACACCCGAAACGCTGATCGATATTGTGTATCAGCCACAGCGGTTTTCAAAGTG 303  
|  
Db 100274 TCAGTGGACACCCGAAACGCTGATCGATATTGTGTATCAGCCACAGCGGTTTTCAAAGTG 100333  
|  
QY 304 CGCCAGTGACAAGATGCACGAGTTCATGCCGGACACGCCAGGCTGTGTTTCGCTG 363  
|  
Db 100334 CGCCAGTGACAAGATGCACGAGTTCATGCCGGACACGCCAGGCTGTGTTTCGCTG 100393  
|  
QY 364 AATTTCAGCCCGGATGTTGCTCATCTCGCCAGTGGGAAGTGGCAGACACAGTCGCGATTG 423  
|  
Db 100394 AATTTCAGCCCGGATGTTGCTCATCTCGCCAGTGGGAAGTGGCAGACACAGTCGCGATTG 100453  
|  
QY 424 TGGGATCTTAACACAGACACCGCACTTCACTGCACAGGTCTAAGCAGTGGGTTCTG 483  
|  
Db 100454 TGGGATCTTAACACAGACACCGCACTTCACTGCACAGGTCTAAGCAGTGGGTTCTG 100513  
|  
QY 484 TCGGTATCCTGGGCTCCGGATGGCAACGCTTGGCCAGCGGTTGCCAAGCGGCTCTATA 543  
|  
Db 100514 TCGGTATCCTGGGCTCCGGATGGCAACGCTTGGCCAGCGGTTGCCAAGCGGCTCTATA 100573  
|  
QY 544 ATCATCTGGACCCCGAGACGGGTCAGCAGAAGGGGCGACCCCTTGAGTGGGCACAAGAAA 603  
|  
Db 100574 ATCATCTGGACCCCGAGACGGGTCAGCAGAAGGGGCGACCCCTTGAGTGGGCACAAGAAA 100633  
|  
QY 604 CACATCACTGCCTCGCTGGGAACCGTATCATCGGATCCGGAGTCAGGAACCTTGCT 663  
|  
Db 100634 CACATCACTGCCTCGCTGGGAACCGTATCATCGGATCCGGAGTCAGGAACCTTGCT 100693  
|  
QY 664 TCCGCCAGTGGAGACGGGACTGCCGGATTGGGAGTTGGGAGTGGGAGCGGCTTATG 723  
|  
Db 100694 TCCGCCAGTGGAGACGGGACTGCCGGATTGGGAGTTGGGAGTGGGAGCGGCTTATG 100753  
|  
QY 724 AACATTGCCGGACACACAAATGCTGTACAGCAGTGAAGATGTGGGAGCGGCTTATG 783  
|  
Db 100754 AACATTGCCGGACACACAAATGCTGTACAGCAGTGAAGATGTGGGAGCGGCTTATG 100813  
|  
QY 784 TATACATCCTCCAAAGATCGCACAGTGAAGATGTGGGAGCAGCTGTGATGAATCTTGTGC 843  
|  
Db 100814 TATACATCCTCCAAAGATCGCACAGTGAAGATGTGGGAGCAGCTGTGATGAATCTTGTGC 100873  
|  
QY 844 CGGACGTTCTCTGGCCAGCTCACTGGGTAAACAACATTCGCTGAGCACCGGATACGTC 903  
|





QY 244 TCAGTGGACACCGAANAACGTCATCGATATTGTGTATCAGCCACAGGCGGTTTTCRAAGTG 303  
Db 89498 TCAGTGGACACCGAANAACGTCATCGATATTGTGTATCAGCCACAGGCGGTTTTCRAAGTG 89557  
QY 304 CGCCAGTGCACAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG 363  
Db 89558 CGCCAGTGCACAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG 89617  
QY 364 AATTTCAGCCCGGATGGTCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCATTTG 423  
Db 89618 AATTTCAGCCCGGATGGTCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCATTTG 89677  
QY 424 TGGATCTTAACACAGAGACACCGGCACCTTCACCTGCACAGGTCATAAGCAGTGGGTCTG 483  
Db 89678 TGGATCTTAACACAGAGACACCGGCACCTTCACCTGCACAGGTCATAAGCAGTGGGTCTG 89737  
QY 484 TGGATCTTAACACAGAGACACCGGCACCTTCACCTGCACAGGTCATAAGCAGTGGGTCTG 543  
Db 89738 TGGATCTTAACACAGAGACACCGGCACCTTCACCTGCACAGGTCATAAGCAGTGGGTCTG 89797  
QY 544 ATCATCTGGACCCGGAGACGGGTCAGCAAGGGGCGACCCCTTGAGTGGGCACAGAAA 603  
Db 89798 ATCATCTGGACCCGGAGACGGGTCAGCAAGGGGCGACCCCTTGAGTGGGCACAGAAA 89857  
QY 604 CACATCAACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAACTGCT 663  
Db 89858 CACATCAACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAACTGCT 89917  
QY 664 TCCCGCAGTGGAGACGGGACTGCCGGATTGGGACGTAANAATTGGCCAGTGCCTTATG 723  
Db 89918 TCCCGCAGTGGAGACGGGACTGCCGGATTGGGACGTAANAATTGGCCAGTGCCTTATG 89977  
QY 724 AACATTGCCGGACACACAAATGCTGTGACACAGTGGAGTGGGTTGGAGCGGCGCTTAT 783  
Db 89978 AACATTGCCGGACACACAAATGCTGTGACACAGTGGAGTGGGTTGGAGCGGCGCTTAT 90037  
QY 784 TATACATCTCCAAAAGATCGCACAGTGAAGATGTGGCGACGAGTGAATCTTGTGC 843  
Db 90038 TATACATCTCCAAAAGATCGCACAGTGAAGATGTGGCGACGAGTGAATCTTGTGC 90097  
QY 844 CGGACGTTCTGTGGCCAAAGTCACTGGGTAACAACATTCGCTGAGACCCGATFACGTC 903  
Db 90098 CGGACGTTCTGTGGCCAAAGTCACTGGGTAACAACATTCGCTGAGACCCGATFACGTC 90157  
QY 904 CTGGCCACTGGTCCATTCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTGAGC 963  
Db 90158 CTGGCCACTGGTCCATTCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTGAGC 90217  
QY 964 A-----C 965  
Db 90218 AGTAAGAATTTCATATGCTCTATATATGCTGTATATCTCTAAAGAGATTTGTTTACAGC 90277  
QY 966 TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCGGTGTGCCCTGACGAGTGGA 1025  
Db 90278 TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCGGTGTGCCCTGACGAGTGGA 90337  
QY 1026 GTCGCTGGTTTCTGTTCCGATGACAAACACCTCTATCTGTGGCGGAACAACCCAGAACAA 1085  
Db 90338 GTCGCTGGTTTCTGTTCCGATGACAAACACCTCTATCTGTGGCGGAACAACCCAGAACAA 90397  
QY 1086 GTGCGTTGAGCGCATGACAGGCGCACCAAGTGGTCAACGATGTGAATATTTCGCGGGA 1145  
Db 90398 GTGCGTTGAGCGCATGACAGGCGCACCAAGTGGTCAACGATGTGAATATTTCGCGGGA 90457  
QY 1146 TGTAAGCTAATGCGTCTGCTTCATTTGACAAAGTCAAGTCCGTCTGTGGCGAGCCAGCGA 1205  
Db 90458 TGTAAGCTAATGCGTCTGCTTCATTTGACAAAGTCAAGTCCGTCTGTGGCGAGCCAGCGA 90517  
QY 1206 TGGTCAGTACATGGCCACCTCCGGGGTCATGTGCAGGCTGTTTACAGGTTGCTTGGTC 1265  
Db 90518 TGGTCAGTACATGGCCACCTCCGGGGTCATGTGCAGGCTGTTTACAGGTTGCTTGGTC 90577  
QY 1266 CGCGGACTCCCGCTTGATTGTTCCGGGCAGCAAGACTCAACTCTAAA----- 1313

Db 90578 CGCGGACTCCCGCTTGATTTTCCGGCAGCAAGACTCAACTCTAAAGGTTAGCACGT 90637  
QY 1314 -----AGTATGGAGTGTGCAGA 1330  
Db 90638 AGTTTCAGTTTCTAGCGCTAGCTAATACATATCTCTTGTAGTATGGAGTGTGCAGA 90697  
QY 1331 CGAAGAAACTGGCACAGGAGCTGCTGGACATCGCGGATGAGGTGTTTCGGAGTGGACTGG 1390  
Db 90698 CGAAGAAACTGGCACAGGAGCTGCTGGACATCGCGGATGAGGTGTTTCGGAGTGGACTGG 90757  
QY 1391 CGCCCGATGGCTCTAGAGTTGCCCTGGTGGCAAGGACAAAGTTATAAAGCT----- 1442  
Db 90758 CGCCCGATGGCTCTAGAGTTGCCCTGGTGGCAAGGACAAAGTTATAAAGCTGTAAGTTT 90817  
QY 1443 -----ATGGGCTTA 1451  
Db 90818 GAGTAACGAGGTTTTCTATAACAATATACAGACTCTAATTGTATCTTTCAGATGGGCTTA 90877  
QY 1452 TTAACAATCATTAATTTGTACACGTAAGAAATACTTAGGAATAAAGTAAACGTCCT 1511  
Db 90878 TTAACAATCATTAATTTGTACACGTAAGAAATACTTAGGAATAAAGTAAACGTCCT 90937  
QY 1512 GAGTAAA 1518  
Db 90938 GAGTAAA 90944

RESULT 7  
AE003589/c  
LOCUS Drosophila melanogaster genomic scaffold 14200001386046 section 2  
DEFINITION of 16, complete sequence.  
ACCESSION AE003589 AE002638  
VERSION AE003589.2 GI:10727410  
KEYWORDS HTG.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 302473)  
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottiier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S., Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,

Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,  
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,  
Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,  
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,  
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,  
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,  
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
The genome sequence of Drosophila melanogaster  
Science 287 (5461), 2185-2195 (2000)  
20196006  
10731132  
2 (bases 1 to 302473)  
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
On Oct 9, 2000 this sequence version replaced gi:7296179.  
Location/Qualifiers  
1. 302473  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
<3618. .>39613  
/gene="Plc21C"  
/note="CG4574"  
/map="21B7-21C1"  
/db\_xref="FLYBASE:FBan0004574"  
/db\_xref="FLYBASE:FBgn0004611"  
join(3618. .3958,11098. .11376,11383. .11475,19033. .19179,  
27475. .27612,28770. .28903,29373. .29558,29704. .29901,  
30371. .30553,31293. .31473,32562. .32796,34065. .34210,  
35933. .36146,36220. .37083,37396. .37843,37909. .38301,  
38363. .38507,38569. .39613)  
/gene="Plc21C"  
/product="CT14752"  
/db\_xref="FLYBASE:FBan0004574"  
/db\_xref="FLYBASE:FBgn0004611"  
join(11383. .11475,19033. .19179,27475. .27612,28770. .28903,  
29373. .29558,29704. .29901,30371. .30553,31293. .31473,  
32562. .32796,34065. .34210,35933. .36146,36220. .37083,  
37396. .37843,37909. .38301,38363. .38507,38569. .38802)  
/gene="Plc21C"  
/note="Plc21C gene product"  
/codon\_start=1  
/protein\_id="AAF51507.1"  
/db\_xref="FLYBASE:FBan0004574"  
/db\_xref="FLYBASE:FBgn0004611"  
/db\_xref="FLYBASE:FBan0004574"  
/db\_xref="FLYBASE:FBgn0004611"  
/translation="MMSAGGTYISTASVEVPQALQDGEKFIKRWDDSGTGTPTVMRVD  
AKGFLVNDQNNELDILDIAIRDTGTQYAKRPDKNLRIQIVLGPQDTLEEKTVT  
VCHGSDFVNTFTVFCCTRRDIAQLWTDGLIKLAYSIAQLNGSAIMFLQKAHTKLCIQ  
VDKSGRIPVKNIIKLFQAKNEDKRKEKALDVTGLPSGKVDISIVSKFQFEDFYLYK  
YLTORSEVERLFDISIVGNSKRCMSIAQLVEFLNKTQDRPLNEILYPYANPARAKEL  
IQYEPNFKNAKGQLSLDGFLRYLMGDDNPIMAPSKLDLDCDDMDQPMSHYFINSSH  
TYLTGHQLTGKSSVEIYRQCLLAGCRCVELDFWNGRTEEPVIVHGYTFVPEIFAKDVL  
EAIAESAFKTSEYPVILSFENHCNPRQAKIANCYCREIFGDMLLDKPLDPSHPLEPNMD  
LPPPAMLRKRIIKNKKHHHHHHHKKPAQVGTGTPAANNKLTITANSVDAKAAQVVG  
LSASHEDGGVTRSTANGDYATGTGTGSAAGTAGHAPPLQIIRQSSKDDTSSSDSSSS  
EDESLPNTPTNLPNGNEPPEKAQKETAGAEISALVNYVQPIHFSFENAEKKNRCY  
EMSSFDEKQATLLKERPIEFVYNNKHQLSRVYPAGTRFDSNEMPQLFWNAGCOLVA  
LNFQTLDAQLNLGIFEYNARSGYLLKPEFMRRSRDRRLDPPAESTVDGIIAGTVSIT  
VLSGQFLTDRANTFVEVDMYGLPADTVRKFKRTKTVRDNGMNPVDEEPFVKVYL  
PELASIRIAAYEEGKLIGHRVLPVIGLCPGYRHNLRSEVGQPIALASFLCVVVKD  
YVPDDLNSFAEALANPIKYQSELEKRDQLSVLTDEAALGSADELSKSGQKKELR  
PVESLATSPKRPSPISAAAMSVDVTVDRTDGGREGDSISIVAPSIHQHSLDQSVST  
SIRQVESSQDVLVLAEPLEKILDHKSVKEKRLKLEMEKKLESRLKKHDKKIKIAGOK  
SSPLEGKKPKFATTNKLVKRLSNKLSNCLSPHSEPGVEIPACPLDLGDSSEESAADA  
GEDLAGGSSSLDGRTOESRLRSACREYTSQYRELOEKYHEAIIYSAAEKVLKTSQTGOT  
KQLKASLDKYTGEVMHQLOEARRNEVKNLATVHRDRDELIRMKREVASSVVERGVAER  
VRLKQTFDRRTDELQKHDSRNALAEHRSAKQIILDKAEASRCSVSGNGFLVLFHGP  
HHGCTGSGSSALSGNNLTNLNDAGAAGSHSAISPAKSHNSIAAAAEKMT"  
complement(<11712. .>12601)  
/gene="CG11912"

/map="21B7-21B8"  
/db\_xref="FLYBASE:FBan0011912"  
/db\_xref="FLYBASE:FBgn0031248"  
complement(11712. .12601)  
/gene="CG11912"  
/product="CT37080"  
/db\_xref="FLYBASE:FBan0011912"  
/db\_xref="FLYBASE:FBgn0031248"  
complement(11773. .12588)  
/gene="CG11912"  
/note="CG11912 gene product"  
/codon\_start=1  
/protein\_id="AAF51509.1"  
/db\_xref="GI:7296217"  
/db\_xref="FLYBASE:FBan0011912"  
/db\_xref="FLYBASE:FBgn0031248"  
/translation="MKQFAVIFALALASVS AISVPOPGFPPEGRIINGVEAAKGEAPYI  
VSLQTTSNSHFCAGSLDEVTIVTAAHCLTYNQQAVAGAHSRTOENQVIRKFTNAQ  
YVIHENYGGVGPNDIGILLKEEDAFDLNAVARDGNSNPVSAVSLPSKTFQGTSDGYL  
YWGWRDNSGLLPLNLQKLDATIVDYNCKAALPSNNSLAETNVCVTHPPGKADGSCNGD  
SGGPLVSSSSRGAELIGIVSWGYPCLSTTYPVYTSVSSFLPWIDENRKAK"  
complement(<13648. .>14580)  
/gene="CG11911"  
/map="21B7-21B8"  
/db\_xref="FLYBASE:FBan0011911"  
/db\_xref="FLYBASE:FBgn0031249"  
complement(13648. .14580)  
/gene="CG11911"  
/product="CT37076"  
/db\_xref="FLYBASE:FBan0011911"  
/db\_xref="FLYBASE:FBgn0031249"  
complement(13743. .14576)  
/gene="CG11911"  
/note="CG11911 gene product"  
/codon\_start=1  
/protein\_id="AAF51508.1"  
/db\_xref="GI:7296216"  
/db\_xref="FLYBASE:FBan0011911"  
/db\_xref="FLYBASE:FBgn0031249"  
/translation="MKLIYTVTLVIALVAAAQGA LSKLAKLVPSFATGCVINGTEAE  
PHSAPYIVSLATNYLKHSHICGGTLINKDWIVTAAHCISEFPVGMSSIIAGLHTRAEVDE  
LTQORQVDFGRVHEKYTGVPYDIALHVNESFIFNEWQPATLPSRQVHEGETHL  
YWGQPKSYIFSGAKTLQVTTQIILNYEKEELPESAPIAESNICSSLSLOQSKSACN  
GDSGGPLVVEFTNAPSELIGIVSWGYPICGLANMPSIYTKVSAYIDWITNIQSAYYKL  
F"  
complement(<49281. .>51230)  
/gene="CG11907"  
/map="21C1-21C1"  
/db\_xref="FLYBASE:FBan0011907"  
/db\_xref="FLYBASE:FBgn0031250"  
complement(join(49281. .49459,49730. .50200,50254. .50554,  
50786. .>51230))  
/gene="CG11907"  
/product="CT33136"  
/db\_xref="FLYBASE:FBan0011907"  
/db\_xref="FLYBASE:FBgn0031250"  
complement(join(49281. .49459,49730. .50200,50254. .50554,  
50786. .51049))  
/gene="CG11907"  
/note="CG11907 gene product"  
/codon\_start=1  
/protein\_id="AAF51506.1"  
/db\_xref="GI:7296214"  
/db\_xref="FLYBASE:FBan0011907"  
/db\_xref="FLYBASE:FBgn0031250"  
/translation="MPGTSDRRLVDALDEDEILGTEGSFVPAPFRPQDHHADEQCLM  
ERQODEVVLVSNPESGRLFTYLVFYLLGIGTMTPNWFFVTAEDYWKYKFRNASINNT  
DLEELTPLOKSFCDLALTATISGTTFLLLNAIFCHVSLRKMGLTLMILILFVG  
TTGFEINTDQWQFFLITLIIIVLLNISAATMSGALYGVAGLFPSEFITAVVSGQA  
LGLILTALAFILVLAFTGPNPTAFIFFIVGGVLLILLCIVCYVILARKPFFRYYLEGG  
DKYKVIKRVPSHNRNGSAEGLPLEPILLRQVMSKIYLHAISLALYTTLSVYPATVVL  
MQSEYGHSVWTAMVMVFALSNGYFTNILLIMAFKRVQKEKELASSIMAAALSCGMV  
GSLLSLVFVQML"



gene	<68682..>77995 /gene="al" /note="CG3935" /map="21C1-21C2" /db_xref="FLYBASE:FBan0003935" /db_xref="FLYBASE:FBgn0000061" join(68682..69037,75067..75450,76257..76302,76864..77132,77259..77995) /gene="al" /product="CTL2963" /note="Nucleotide sequence of the Celera sequence differs			
Query Match	82.7%;	Score 1285.8;	DB 3;	Length 302473;
Best Local Similarity	89.2%;	Pred. No. 6.6e-311;		
Matches 1505;	Conservative 0;	Mismatches 7;	Indels 175;	Gaps 4;
QY	7	CAAAAAATGCAGGAGACGGACACACGGAGCAAGAGGCCACGCCACATACGATACACAGGCGCGC	66	
Db	258606	CAGAAAATGCAGGAGACGGACACACGGAGCAAGAGGCCACGCCACATACGATACACAGGCGCGC	258547	
QY	67	CTCGTTT---ACACGGCGAGGAAGCCGGCCCGCCCAATCGACCTGCCGGCAGGAATCACT	123	
Db	258546	CTCGTTTCGGACACGGCGGAGGAAGCGGGCCCGCCCAATCGACCTGCCGGCAGGAATCACT	258487	
QY	124	ACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT	183	
Db	258486	ACCCAGCAATTGGGTCTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT	258427	
QY	184	TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG	243	
Db	258426	TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG	258367	
QY	244	TCAGTGGACACCGAAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTG	303	
Db	258366	TCAGTGGACACCGAAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTG	258307	
QY	304	CGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG	363	
Db	258306	CGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG	258247	
QY	364	AATTCAGCCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGGGATTG	423	
Db	258246	AATTCAGCCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGGGATTG	258187	
QY	424	TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCAGTGGGTTCTG	483	
Db	258186	TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCAGTGGGTTCTG	258127	
QY	484	TGCGTATCCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAGCGGGCTCTATA	543	
Db	258126	TGCGTATCCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAGCGGGCTCTATA	258067	
QY	544	ATCATCTGGGACCCGGAGACGGGTACGAGAAGGGCGGACCCTTGAGTGGGCACAAGAAA	603	
Db	258066	ATCATCTGGGACCCGGAGACGGGTACGAGAAGGGCGGACCCTTGAGTGGGCACAAGAAA	258007	
QY	604	CACATCAACTGCCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGCT	663	
Db	258006	CACATCAACTGCCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGCT	257947	
QY	664	TCCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAATAATTGGGCCAGTGCCTTATG	723	
Db	257946	TCCGCCAGTGGAGACGGGACTGCCGGATTTGGGACGTAATAATTGGGCCAGTGCCTTATG	257887	
QY	724	AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGGTGGAGCGGCTTATT	783	
Db	257886	AACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGGTGGAGCGGCTTATT	257827	
QY	784	TATACATCCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC	843	
Db	257826	TATACATCCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC	257767	
QY	844	CGGACGTTCTCTGGCCCAAGCTCACTGGGTAACACACATTGCCGCTGAGCACCGGATTACGTC	903	

Db	257766	CGGACGTTCTCTGGCCACGCTCACTGGGTAAACAACATTCGCTGAGCACCGACTACGTC	257707	
QY	904	CTGCGCACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC	963	
Db	257706	CTGCGCACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCCACCTCAGTTTGAGC	257647	
QY	964	A-----C	965	
Db	257646	AGTAAGAAATTCATATGCTCTCTATATATGCTGTATATCCTAAAAGAGATTGTTTACAGC	257587	
QY	966	TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAAGCGCTGTGCCCTGACGAGGTGGA	1025	
Db	257586	TGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAAGCGCTGTGCCCTGACGAGGTGGA	257527	
QY	1026	GTGCGTGGTTTCCTGTTTCGGATGACAACACCCCTCTATCTGTGGCGGAACAACAGAACAA	1085	
Db	257526	GTGCGTGGTTTCCTGTTTCGGATGACAACACCCCTCTATCTGTGGCGGAACAACAGAACAA	257467	
QY	1086	GTGCGTTGAGCGCATGACAGGSCACCAGAACCTGGTCAACGATGTGAAATATTCGCCGGA	1145	
Db	257466	GTGCGTTGAGCGCATGACAGGSCACCAGAACCTGGTCAACGACGTGAAATATTCGCCGGA	257407	
QY	1146	TGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTGCCTGTGTGGCGAGCCAGCGA	1205	
Db	257406	TGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTGCCTGTGTGGCGAGCCAGCGA	257347	
QY	1206	TGTCAGTACATGGCCACCTTCCGGGTCATGTGCAGGCTGTTTACACGGTTCCTTGGTC	1265	
Db	257346	TGTCAGTACATGGCCACCTTCCGGGTCATGTGCAGGCTGTTTACACGGTTCCTTGGTC	257287	
QY	1266	CGCGGACTCCCCTTGATTGTTTCCGGCAGCAAAAGACTCAACTCTAAAAGTTAGCACGT	1313	
Db	257286	CGCGGACTCCCCTTGATTGTTTCCGGCAGCAAAAGACTCAACTCTAAAAGTTAGCACGT	257227	
QY	1314	-----AGTATGGAGTGTGCAGA	1330	
Db	257226	AGTTTCAGTTTCTAGCGCTAGCTAATACATATCCTCTTCTTCTAGTATGGAGTGTGCAGA	257167	
QY	1331	CGAAGAACTGGCACAGGAGCTGCCTGGACATCGGATGAGGTGTTCCGGAGTGGACTGGG	1390	
Db	257166	CGAAGAACTGGCACAGGAGCTGCCTGGACATCGGATGAGGTGTTCCGGAGTGGACTGGG	257107	
QY	1391	CGCCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGACAAAGTTATAAAGCT-----	1442	
Db	257106	CGCCCGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGACAAAGTTATAAAGCTGTAAGTTT	257047	
QY	1443	-----ATGGGCTTA	1451	
Db	257046	GAGTAACGAGGTTTTCTATAACAATATACAGACTCTAATTGTATCTTTCAGATGGGCTTA	256987	
QY	1452	TTAACAAATCATTAACCTTGTACACGGTAAGAAAACTACTAGGAATAAAGTAAACGTCCT	1511	
Db	256986	TTAACAAATCATTAACCTTGTACACGGTAAGAAAACTACTAGGAATAAAGTAAACGTCCT	256927	
QY	1512	GAGTAAA	1518	
Db	256926	GAGTAAA	256920	
RESULT 8				
AX135811				
LOCUS	AX135811	1859 bp	DNA	linear
DEFINITION	Sequence 5 from Patent WO0132614.			
ACCESSION	AX135811			
VERSION	AX135811.1	GI:14272046		
KEYWORDS	human.			
SOURCE	ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 1859)			
AUTHORS	Utku,N.			



TITLE Novel genes tzap7/a, tzap7/b and tzap7 involved in t cell activation and uses thereof  
JOURNAL Patent: WO 0132614-A 5 10-MAY-2001;  
Utku, Nalan (DE)  
FEATURES Location/Qualifiers  
source 1..1859  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
CDS 20..1477  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC39576.1"  
/db\_xref="GI:14272047"  
QLVCNALLAQEDPLPLAFFVHDAEIVSSLGKLTLESQAVETEKVLDIYQPOAIFRVA  
VTRCTSSLEGHSEAVISVAFSPTKYLASGSGDITVRFWDLSTETPHFTCKGHRHWVL  
SISWSPDGKKLASGCKNGQILLWDPSPGKQVGRITLAGHSKWITGLSWEPLHANPECRY  
VASSSKDGSVRIWDTTAGRCERILTGHTQSVTCLRWGDDGLLYSASODRTIKVYRAHD  
GVLCRTLQGHGHWNMTMALSTDYALRTGAFEPAAESVNPQDLQSGSLQELKERALSRYN  
LVRGQGPRLVSGSDDEFTFLWSPAEDKKPLTRMTGHOALINQVXFPSPDSRIVASASF  
DKSIKLWDGRTGKYLASLRGHVAAVYQIAWSADSRLLVSGSSDSTLKVWDVKAQKLAM  
DLPGHAEVYAVDWSPDQGVASGGDKCLRWR"

BASE COUNT 392 a 545 c 566 g 355 t 1 others  
ORIGIN  
Query Match 30.7%; Score 478; DB 6; Length 1859;  
Best Local Similarity 59.8%; Pred. No. 8.5e-109;  
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;  
QY 54 GATACAGGCGCGCTCGTTTACACGGGCGAGGAAGCCGGCCGCAATCGACCTGCCCGC 113  
Db 70 GCTAGTGCAGTTCACGATGAGGGCGGCAGCTGCTGGGTTCGCCGTTCGACGTGCCCGT 129  
QY 114 AGGAATCACTACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAACACGAGGAAGC 173  
Db 130 GGACATCACCCCGGACAGGCTGCAGCTCGTGTGCAACGCGCTACTGGCCCCAGGAGATCC 189  
QY 174 CACTCCATATTTGTTTTCGTGGCGGAGGATGAGATCAAGAAGACCTGGAGGACACGTT 233  
Db 190 CCTGCCACTGGCTTTCTTGTCCACGATGCTGAGATGCTCTCCTCACTGGGGAACACGTT 249  
QY 234 GGACTTGGCGTCAGTGGACACCGAAACGTCATCGATATTTGTATATCAGCCACAGGCGGT 293  
Db 250 GGAGTCCAGGCAGTGGAGACAGAGAAGGTCCTAGACATCATCTACCAGCCACAGGCTAT 309  
QY 294 TTTCAAAGTCCGCCAGTGACAAGATGCACGAGTTCATATCCCGGGGACACGCCGAGGCTGT 353  
Db 310 CTTCAGAGTCCGGCTGTGACTCGCTCCACCAGCTCCTTGGAGGTCACAGTGAGGCAGT 369  
QY 354 GGTTCGCTGAATTTACGCCCCGGATGGTGTCTCATCTGCCAGTGGGAAGTGGCGACACCAC 413  
Db 370 CATTTCTGTGGCCTTCAGCCCTACGGGAAAGTACCTTGGCCAGTGGCTCTGGAGACACCAC 429  
QY 414 AGTCCGATTGTGGGATCTTAACACAGACACACCGCACTTCACCTGCACAGGTCATAAGCA 473  
Db 430 CGTGGCCTTCTGGGATCTCAGACACAGACACACCACATTTACATGCAAGGGACACACAGACA 489  
QY 474 GTGGGTTCTGTGCGTATCCTGGGCTCCGGATGGCAACAGGTTGGCCAGCGGTTGCAAAAGC 533  
Db 490 CTGGGTCCTTAGTATATCCTGGTCTCCAGATGGCAAGAGCTGSCCTCAGGCTGCAAGAA 549  
QY 534 GGGCTCTATAATCATCTGGGACCCGGAGACGGGTTCAGCAGAAGGGGGACCCCTTGAGTGG 593  
Db 550 TGGCCAGATTCTCCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGGCGAGGACCCCTCGCTGG 609  
QY 594 GCACAAGAAACACATCAACTGCCTCGCTGGGAACCGTATCATCGCGATCCGGAGTGCAG 653  
Db 610 CCACAGCAAGTGGATCACAGGCGCTGAGCTGGGAGCCCTCCATGCGGAACCCCTGAGTGCCG 669  
QY 654 GAAACTTGCTCCGCCAGTGGAGACGGGGAAGTGGCGGATTTGGGACGTAAATTTGGGCCA 713  
Db 670 CTATGTGGCCAGCAGCTCCAAGGATGGCAGTGTGCGGATCTGGGACACAACTGCAGGCCG 729

QY 714 GTGCCATTATGAACATTGCCGGACACACAAATGCTGTGACACAGCAGTGGGTGGAGC 773  
Db 730 CTGTGAGCGCATCCTCACCAGGACACACCCAGTGGTACCTGTCTCCGGTGGGAGGGA 789  
QY 774 GGGCCTTATTATACATCCTCCAAAGATCGCACAGTGAAGATGTGGGAGCAGCTGATGG 833  
Db 790 CGGGCTTCTACTCTCGCTCCCAGGACCGCACCATCAAGTCTGGAGAGCTCATGACGG 849  
QY 834 AATCTTGTGCCGGACGTTCTCTGGCCRAAGCTCACTGGGTAAACAACATTCGCTGAGCAC 893  
Db 850 TGTGCTGTGCCGACTCTGCAAGGGCCACGGCCACTGGTGAACACCACTGGCCCTCAGCAC 909  
QY 894 CGATTACGTCCTGCGCACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCC--A 950  
Db 910 TGACTATGCCCTGCGCACTGGGGCCCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCCAAGA 969  
QY 951 CCTCAGTTTGAGCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAAGCCGCTGTG 1010  
Db 970 CCTCCAAGGATCCTTGCAGGAGTTGAAGGAGAGGGCTCTGAGCCGATACAAACCTCGTGG 1029  
QY 1011 CCCTGACGAGGTGGAGTCTGCTGGTTCTTCTGTTTCGGATGACAAACACCCCTCTATCTGTGG-- 1068  
Db 1030 GGGCCAGGCTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACCTTATCTGTGGTCT 1089  
QY 1069 -CGGAACACACAGAACAGTGCCTTGAGCGCATGACAGGSCACACAGACGTTGGTCAACGA 1127  
Db 1090 CCCAGCAGAGGACAAAAGCCTCTCAGTCGGATGACAGACACCAAGCTCTCATCAACCA 1149  
QY 1128 TGTGAAATATTGCCCGGATGTAAAGCTAAATGCGTCTGCTTCATTTGACAAGTCAAGTGG 1187  
Db 1150 GGTGNTCTTCTCTCTGACTCCCGCATCTGCTGGCTAGTGCTCTCTTTGACAAGTCCATCAA 1209  
QY 1188 TCTGTGGGAGCCAGCGATGGTCAGTACATGCGCCACCTTCCGGGGTCACTGTGAGGCTGT 1247  
Db 1210 GCTGTGGGATGGCAGGACGGCAAGTACCTGGCTTCCCTACGCGGCCACGTGGTGGCT 1269  
QY 1248 TTACACGGTTGCTTGGTCCGCGGACTCCCGCTTGATTGTTTCCGCGACGAAGACTCAAC 1307  
Db 1270 GTACCAGATTGCGTGGTCACTGACAGTCCGCTCCTGGTCAAGCGGACGAGTGCAGCAC 1329  
QY 1308 TCTAAAAGTATGGAGTGTGCAGACGAAGAACTGGCACAGGAGCTGCCTGGACATGCCGA 1367  
Db 1330 ACTGAAGGTGTGGGATGTGAAGGCCCAAGAGCTGGCCATGGACCTGCCCGGCCACGCGGA 1389  
QY 1368 TGAGGTGTTCCGAGTGGACTGGCGGCCCGGATGGCTCTAGAGTTGCCTCTGGTGGCAAGGA 1427  
Db 1390 TGAGGTATATGCTGTTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA 1449  
QY 1428 CAAAGTTATAAAGCTATGG 1446  
Db 1450 CAAATGCCCTCCGGATATGG 1468

RESULT 9  
AK001320  
LOCUS Homo sapiens cDNA FLJ10458 fis, clone NT2RP1001457, highly similar to Homo sapiens partial mRNA for beta-transducin family protein.  
DEFINITION AK001320  
ACCESSION AK001320  
VERSION AK001320.1 GI:7022501  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens teratocarcinoma cell line:NT2 cDNA to mRNA, clone\_lib:NT2RP1 clone:NT2RP1001457.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wajitsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,







COMMENT

Contact: MGC help desk  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,  
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 15 Row: g Column: 15  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7022501.  
Location/Qualifiers

FEATURES

source

1..1870  
/organism="Homo sapiens"  
/db\_xref="LocusID:54475"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3943492"  
/tissue\_type="Lung, small cell carcinoma"  
/clone\_lib="NIH\_MGC\_7"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
<1..1465  
/codon\_start=2  
/product="hypothetical protein similar to beta-transducin  
family"  
/protein\_id="AAH02884.1"  
/db\_xref="GI:12804063"  
/translation="ARGAAVADEAVARDVQRLLVQFDEGGQLLGSFPDVPVDTIPD  
RLQLVCNALLAQEDPLPLAFFVHDAEIVSSLGKTLSEQAVETEKVLDIYQPQALFRV  
RAVTRCTSLSEGHSEAVISVAFSPTGKYLASGSDITVRFWDLSTETPHFTCKGHRHW  
VLSISWSPDGKKLASGCKNGQILLDPSTGKQVGRTLAGHSKWIITGLSWEPLHANPEC  
RYVASSKDGSVRIWDTAGRCERILFGHTQSVTCLRWGGDGLLYSASQDRITIKYWR  
HDGVLCTRLOGHGWNTMALSTDYALRTGAFPAEASVNPQDLQSLQELKERALS  
YNLVRGQPERLVSGSDFTFLWSPAEKKPLTRMTGHQALINQVLFSPDSRIYASA  
SFDKSIKLWDGRTGKYLASLRGHVAAYQIAWSADSLRLVSGSDSLKLVWDVYKAKL  
AMDLPGHADVEYAVDWSPDGQVRVASGGKDKCLRWR"

CDS

BASE COUNT  
ORIGIN

Query Match	30.7%;	Score	477.4;	DB	9;	Length	1870;
Best Local Similarity	59.8%;	Pred.	No. 1.2e-108;				
Matches	837;	Conservative	0;	Mismatches	556;	Indels	6;
Gaps	2;						
QY	54	GATACAGCGCGCCTCGTTTACACGGGGGAGGAGCGCGCCGCCAATCGACCTGCCGCG	113				
Db	58	GCTAGTGCAGTTCAGGATGAGGGCGGCAGCTGCTGGGTTCCCCGTTCCGAGTGCOCGT	117				
QY	114	AGGAATCACTACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGC	173				
Db	118	GGACATCACCCCGGACAGGCTGCAGCTCGTGTGCAACGCGCTACTGGCCCGCAGGAGTCC	177				
QY	174	CACATCCATATTTGTTTTCGTGGGCGAGGATGAGATCAAGAGAGCCCTGGAGGACACGTT	233				
Db	178	CCTGCCACTGGCTTCTTTGTCCACGATGCTGAGATCGTCTCTCTCACTGGGGAACGTT	237				
QY	234	GGACTTGGCGTCACTGGACACCCGAAAAACGTGATCGATATTGTGTATCAGCCACAGCGGT	293				
Db	238	GGACTCCAGGCAGTGGAGACAGAGAAGTCTCTAGACATCATCTACCAGCCACAGGCTAT	297				
QY	294	TTTCAAAGTGCGCCAGTGCACAAGATGCACAGTTCATGCCGGGACACGCCGAGGCTGT	353				

Db	298	CTTCAGAGTCCGGGCTGTGACTCGCTGCACCAGCTCCTTTGGAGGGTCACAGTGAGGCAGT	357				
QY	354	GGTTTCGCTGAATTTACGCCCGGATGGTGTCTCATCTCGCCAGTGAAGTGGGACACACC	413				
Db	358	CATTTCTGTGGCCTTCAGCCCTACGGGAAAGTACCTGGCAGTGGCTCTGGAGACACCAC	417				
QY	414	AGTCCGATTGTGGGATCTTAACACAGACAGACACCGGCACCTTCACCTGCACAGGTCTAA	473				
Db	418	CGTCCGCTTCTGGGATCTCAGCACAGACACACACATTTACATGCAAGGACACAGACA	477				
QY	474	GTGGGTTCTGTCCGTATCCTGGGGCTCCGGATGGGCAACCGTTGGCCAGCGGTGCAAGC	533				
Db	478	CTGGGTCCITAGTATATCCTGGTCTCCAGATGGCAAGAGCTGGCCTCAGGCTGCAAGAA	537				
QY	534	GGGCTCTATAATCATCTGGGACCCCGGAGACGGGTACAGAGAAGGGCGCACCTTGAGTGG	593				
Db	538	TGGCCAGATTCTCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGGCAGGACCCCTCGCTGG	597				
QY	594	GCACAAAGAAACACATCAACTGCCTCGCCCTGGGAACCCGTATCATCGCGATCCGGAGTGCAG	653				
Db	598	CCACAGCAAGTGGATCACAGGCCTGAGCTGGGAGCCCTCCATGCGAACCCCTGAGTGC	657				
QY	654	GAACCTTGCTTCCGCCACTGGAGACGGGACTGCCGGATTGGGACGTAATAATTGGGSCA	713				
Db	658	CTATGTGGCCAGCAGCTCCAAGGATGGCAGTGTGCGGATCTGGGACACAACTGCAGGCGG	717				
QY	714	GTGCCTTATGAACATTGCGGGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGC	773				
Db	718	CTGTGAGCGCATCTCACGGGACACACCCAGTGGTCCCTGTCTCCGTTGGGAGGGGA	777				
QY	774	GGGCTTATTTATACATCCTCCAAAGATCGCACAGTGAAGATGTGGCAGCAGCTGATGG	833				
Db	778	CGGGCTTCTACTCTGCTCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG	837				
QY	834	AATCTTGTGCCGGAGCTTCTTGCCCAAGCTCACTGGGTAAAACAACATTCGCTGAGCAC	893				
Db	838	TGTGCTGTGCCGGACTCTGCAAGGCCACGGCCACTGGGTGAACACCATGGCCCTCAGCAC	897				
QY	894	CGATTACGCTCTGCGCACTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCC--A	950				
Db	898	TGACTATGCCCTCGGCACCTGGGCGCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCAAGA	957				
QY	951	CCTCAGTTTGAGCAGTGAAGAAATTGACGGAATCTGCTTGAAGCGGTACCAGGCCGTGTG	1010				
Db	958	CCTCAAGGATCCTTGCAGGAGTTGAAGGAGAGGGCTCTGAGCCGATACAACTCGTGGC	1017				
QY	1011	CCCTGACGAGGTGGAGTCGCTGGTTTCTGTTCCGATGACAAACACCTCTATCTGTGG--	1068				
Db	1018	GGGCCAGGGTCCAGAGAGGCTGGTGTGGTCCGACGACTTCACTTATTCCTGTGGTC	1077				
QY	1069	-CGGAACAACACAGAACAGTGCCTTGAAGCGCATGACAGGGCACACAGAACTGGTCAACGA	1127				
Db	1078	CCCAGCAGAGGACAAAAGCCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA	1137				
QY	1128	TGTGAATATTCGCCGGATGTAAAGCTAATTGCGTCTGCTTCAATTGACAAAGTCAAGTGGC	1187				
Db	1138	GGTGTCTTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCCITTGACAAAGTCCATCAA	1197				
QY	1188	TCGTGGCGAGCCAGCGATGGTCACTACATGGCCACCTTCCGGGGTTCATGTGCAGGCTGT	1247				
Db	1198	GCTGTGGATGGCAGGACGGGCAAGTACCTGGCTTCCCTACGCGGCCACGTTGGTGCCT	1257				
QY	1248	TTACACGGTTGCCCTGGTCCGCGGACTCCCGCTTCCGCTTGATTGTTTCCGGCAGCAAGACTCAAC	1307				
Db	1258	GTACCAGATTGCGTGTGCTCAGCTGACAGTCGGCTCCTGGTCAGCGGCAGCAGTGACAGCAC	1317				
QY	1308	TCTAAAGATATGGAGTGTGCAGACGAAAGAACTGGCACAGGAGCTGCCTGGACATGCCGA	1367				
Db	1318	ACTGAAGGTGTGGGATGTGAAGGCCCAAGAGCTGGCCATGGACCTGCCCCGCCACGCGGA	1377				
QY	1368	TGAGTGTTCGGAGTGGACTGGGCGCCCGGATGGCTCTAGAGTTGCCCTCTGTGGTGAAGGA	1427				
Db	1378	TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGTGGGAAGGA	1437				



QY 1428 CAAAGTTATAAAGCTATGG 1446  
||||| | | | | |  
Db 1438 CAAATGCCTCCGGATATGG 1456

RESULT 12  
BC012075 2593 bp mRNA linear PRI 06-AUG-2001  
LOCUS Homo sapiens, hypothetical protein similar to beta-transducin  
DEFINITION family, clone MGC:19970 IMAGE:4652713, mRNA, complete cds.  
ACCESSION BC012075  
VERSION BC012075.1 GI:15082334  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2593)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Snailus, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 29. Row: O Column: 21  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7022501.  
Location/Qualifiers  
1. .2593  
/organism="Homo sapiens"  
/db\_xref="LocusID:54475"  
/db\_xref="taxon:9606"  
/clone="MGC:19970 IMAGE:4652713"  
/tissue\_type="Placenta, choriocarcinoma"  
/clone\_lib="NIH\_MGC\_21"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
14. .1471  
/codon\_start=1  
/product="hypothetical protein similar to beta-transducin  
family"  
/protein\_id="AAH12075.1"  
/db\_xref="GI:15082335"  
/translation="MAAAVADEAVARDVQRLLLVQFDEGGQLLGSPDPVDPDITPDRL  
QLVCNALLAQEDPLPLAFFVHDAEIVSSLGKTLSEQAVETEKVLDIYQPQAIFRVRA  
VTRCTSSLEHSEAVISVAFSPTGKYLASGSGDITVRFWDLSTETPHTCKGHRHWVL  
SISWSPDGKKIASGCKNGOILLWDPSTGKQVRTLGHASKWITGLSWEPLHANPECRY  
VASSSKDGSVRIWDTTAGCERILTGHQSVTCLRWGGDGLLYSASQDRTIKVWRAHD  
GVLCRTLQHGHWNTMALSTDYALRTGAFEPAEASVNPQDLQGSLOELKERLSRYN  
LVRGQGPRLVSGSDDFTLFLWSPAEDKKPLTRMTGHQALINQVLFSPDSRIVASAF  
DKSIKLWDGRGKYLASLRGHVAAYQIAWSADSRLLVSGSSDSTLKYMDVKAQKLAM  
DLPGHAEVYAVDWSPDGORVASGGKDKCLRWR"

FEATURES  
source

CDS

BASE COUNT	582 a	734 c	749 g	528 t
ORIGIN				
Query Match	30.7%; Score 477.4; DB 9; Length 2593;			
Best Local Similarity	59.8%; Pred. No. 1.3e-108;			
Matches 837; Conservative	0; Mismatches 556; Indels 6; Gaps 2;			
QY	54	GATACAGGCGCGCCTCGTTTACACGGCGGAGGAAGCGCGCCGCAATCGACCTGCCGCG	113	
Db	64	GCTAGTGCAGTTCACAGATGAGGGCGGCAGCTGCTGGGTTCCCGGTCGACGTGCCCGT	123	
QY	114	AGGAATCAGTACCCAGCAATTGGGACTGATTTGCAACCGCGCTGCTGAAAAACGAGGAAGC	173	
Db	124	GGACATCACCCCGGACAGGCTGCAGCTGCTGTGCAACGCGCTACTTGGCCAGGAGGATCC	183	
QY	174	CACCTCCATATTTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCGCTGGAGGACAGTT	233	
Db	184	CCTGCCACTGGCTTCTTGTCCACGATGCTGAGATCGTCTCCTCACTGGGGAAGAGCTT	243	
QY	234	GGACTTGGCGTCAGTGGACACCCGAAACGTCGATCGATATTGTATCATCAGCCACAGCGGT	293	
Db	244	GGAGTCCAGGCAGTGGAGACAGAGAAGGTCCTAGACATCATCTACAGCCACAGGCTAT	303	
QY	294	TTTCAAAGTGGCCCGCAGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGTGT	353	
Db	304	CTTCAGAGTCCGGGCTGTGACTCGCTGCACCCAGCTCCTTGAGGGTTCACAGTGAGGCAGT	363	
QY	354	GGTTTCGCTGAATTTACGCCCGGATGGTGTCTCATCTGCCAGTGGGAAGTGGCGACACCAC	413	
Db	364	CATTTCTGTGGCCTTCAGCCCTACGGGAAAAGTACCTGGCCAGTGGCTCTTGAGACACCAC	423	
QY	414	AGTGCAGTTGTGGATCTTAACACAGAGACACCGCACTTCACTGACAGGTCATAGACA	473	
Db	424	CGTGGCTTCTGGGATCTCAGCACAGAGACACCACATTTACATGCAAGGGACACACACA	483	
QY	474	GTGGTTCGTGCGTATCCTGGGCTCCGGATGGCAACCGGTTGGCCAGCGGTTGCRAAGC	533	
Db	484	CTGGGTCTTAGTATATCTGGTCTCCAGATGGCAAGAGCTGGCCCTCAGGCTGCAAGAA	543	
QY	534	GGGCTCTATAATCATCTGGACCCCGAGAGCGGTCAGCAGAGGGGCGACCCCTTGAGTGG	593	
Db	544	TGGCCAGATTCTCCTCTGGGACCCCAAGCACAGGGAAGCAGGTTGGGCAGGACCCCTCGCTGG	603	
QY	594	GCACAAGAAACACATCAACTCGCTCGCCTGGGAAACCGTATCATCGCGATCCGGAGTGCAG	653	
Db	604	CCACAGCAAGTGGATCACAGGCGCTGAGCTGGGAGCCCTCCATGCGAACCCTGAGTGCCG	663	
QY	654	GAACACTTGTCCGCCAGTGGAGACGGGACTGCCGGATTGGGAGTAAATAATTGGGCCA	713	
Db	664	CTATGTGGCCAGCAGCTCCAAGGATGGCAGTGTGCGGATCTGGGACACAACAATGCAGGCGG	723	
QY	714	GTGCCTTATGAACATTCGCGGACACACAATGCTGTGACAGCAGTGAAGATGGGTTGGAGC	773	
Db	724	CTGTGAGCGCATCCTCACCGGGCACACCCAGTCGGTCACTGTCTCCGTTGGGAGGGA	783	
QY	774	GGGCTTATTATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG	833	
Db	784	CGGGCTTCTCTACTCTGCCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG	843	
QY	834	AATCTTGTGCGGACGTTCTCTGGCCAAGCTCACTGGGTAAACAACAATTCGCTGAGCAC	893	
Db	844	TGTGCTGTGCGGACTCTGCAAGGGCCACGGCCACTGGGTGAACACCACTGGCCCTCAGCAC	903	
QY	894	CGATTACGTCCTCGCACGTGTCATTCCTCCGTTCCGTTGAAGGATCGCTCCAAGAGCC---	950	
Db	904	TGACTATGCCTCGCACCTGGGGCCCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCCAAGA	963	
QY	951	CCTCAGTTTGAGCACTGAGGAATTGCAGGAATCTGCCCTTGAAGCGCTACACGGCCGTGTG	1010	
Db	964	CCTCCAAGATFCCTTGCAGGAGTTGAAGGAGAGGGCTCTGAGCCCGATACAAACCTCGTGG	1023	
QY	1011	CCCTGACGAGGTGGAGTCGCTGGTTTCTCTGTTCGGATGACAACACCCCTCTATCTGTGG--	1068	

Db 1024 GGGCCAGGGTCCAGAGAGGCTGGTCTCTGGCTCCGACGACTTCACCTTATTCTGTGGTC 1083

QY 1069 -CGGAACAACAGACAAGTGCCTTGAGCGCATGACAGGGCACAGAACGTTGTCACAGA 1127

Db 1084 CCAGCAGAGGACAAAAAGCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA 1143

QY 1128 TGTGAATATTCCCGGATGTAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTGG 1187

Db 1144 GGTGCTCTTCTCTCTGACTCCCGCATCGTGGCTAGTGCCCTCTTGACAAGTCCATCAA 1203

QY 1188 TCTGTGGCGAGCCAGGATGGTCAGTACATGGCCACCTTCCGGGTCATGTGAGGCTGT 1247

Db 1204 GCTGTGGGATGGCAGGACGGCAAGTACCTGGCTTCCCTACGCGGCCACGTGGCTGCCGT 1263

QY 1248 TTACACGGTTGCCGTGGTCCGCGACTCCCGCTTGATTTGTTCCGGCAGCAAGACTCAAC 1307

Db 1264 GTACCAGATTGCGTGGTCAGCTGACAGTCGGCTCTGTCAGCGGCAGCAGTGACAGCAC 1323

QY 1308 TCTAAAAGTATGGAGTGTGCAGAGCAAGAACTGGCACAGGAGCTGCCCTGGACATCGGA 1367

Db 1324 ACTGAAGGTGTGGGATGTGAAGGCCCAAGAGCTGGCCATGGACCTGCCCGGCACGCGGA 1383

QY 1368 TGAGGTGTTCCGAGTGGACTGGCGCCCGCATGGCTCTAGAGTGCCTCTGTGCGCAAGGA 1427

Db 1384 TGAGGTATATGCTGTTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA 1443

QY 1428 CAAAGTTATAAGCTATGG 1446

Db 1444 CAAATGCCCTCCGGATATGG 1462

RESULT 13

AX405715 2638 bp DNA linear PAT 14-JUN-2002

LOCUS Sequence 130 from Patent WO0222660.

DEFINITION AX405715

ACCESSION AX405715

VERSION AX405715.1 GI:21438869

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1

Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A., Ren,F., Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.

TITLE Novel nucleic acids and polypeptides

JOURNAL Patent: WO 0222660-A 130 21-MAR-2002; HYSEQ, INC. (US)

FEATURES

source Location/Qualifiers

1. .2638

/organism="Homo sapiens"

/db\_xref="taxon:9606"

82. .1539

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAD34822.1"

/db\_xref="GI:21438870"

/translation="MAAAVADEAVARDVQRLLVQFQDEGGOLLGSPFVDPVDITPDRL QLVCNALLAQEDPLPLAFFVHDAEIVSSLGKTLSEQAVETEKVLDIIYQPAIFRVRA VTRCTSSLEGHSEAVISVAFSPITGKYLASGSDTTVRFDLSTETPHFTCKGHRHWL S1SWSPDGKKLASGCKNGQIILLWDPSTGKQVGTLAGHSKWITGLSWEPLHANPECRY VASSSKDGSVRIWDTTAGRCERILTGHQSVTCLRWGGDGLLYSASQDRTIKVWRAHD GVLCRTLQGHGWNTMALSTDYALRTGAFEPAEASVNPDLQGSLOELKERALSRYN LVRGQPERLVSGDDFTLFLWSPAEDKKPLTRMTGHQALINQVLFSPSRIVASASF DKS1KLDWGRTKYLAASLRHVAAYVQIAWSADSRLLVSGSSDSTLKVWDVKAQKLAM DLPGHADVEYAVDWSPDQQRVASGGKDKCLRWR"

BASE COUNT 566 a 767 c 764 g 541 t

ORIGIN

Query Match 30.7%; Score 477.4; DB 6; Length 2638;

Best Local Similarity 59.8%; Pred. No. 1.3e-108;

Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;

QY 54 GATACAGGGCGGCCCTCGTTTACAGGGCGAGGAAGCGCGGCCCGCCCAATCGACCTGCGCGC 113

Db 132 GCTAGTGCAGTTCCAGGATGAGGGCGGCAGCTGCTGGGTTCGCCGTTCGACGTGCCGT 191

QY 114 AGGAATCACTACCCAGCAATTTGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGC 173

Db 192 GGACATACCCCGGACAGGCTGCAGCTGCTGTCACACGCGCTACTGGCCCCAGGAGATCC 251

QY 174 CACTCCATATTTGTTTTCGTGGCGAGGATGAGATCAAGAAGAGCTGGAGGACACGTT 233

Db 252 CCTGCCACTGGCTTCTTTGTCCACGATGCTGAGATCGTCTCCTCACTGGGAACAGCTT 311

QY 234 GGACTTGGCGTCAGTGGACACCCGAAACGTCATCGATATTGTGTATCAGCCACAGCGGT 293

Db 312 GGAGTCCAGGTCAGTGGAGACAGAGAAGTCTTAGACATCATCTACCAGCCACAGGCTAT 371

QY 294 TTTCAAAGTGCGCCCTAGTACAAGATGCACGAGTTCCTATGCCGGGACACGCCAGGCTGT 353

Db 372 CTTCAGAGTCCGGGCTGTGACTCGCTGCACACGCTCCTTGGAGGTCACAGTAGGCAGT 431

QY 354 GGTTCGCTGAATTTAGCCCGGATGGTGTCTCATCTGCCAGTGAAGTGGCGACACCAC 413

Db 432 CATTTCTGTGGCTTTCAGCCCTACGGGAAAGTACTTGGCCAGTGGCTCTGGAGACACCAC 491

QY 414 AGTGCATGTGGGATCTTAACACAGACACACCGCACTTCACCTGCACAGGTCATAAGCA 473

Db 492 CGTGCCTTCTGGGATCTCAGCACAGACACACCATTTACATGCAAGGGACACAGACA 551

QY 474 GTGGTCTGTGCGTATCCTGGGCTCCGGATGGCAAAAGGTTGGCCAGCGGTTGCAAGC 533

Db 552 CTGGTCTTAGTATATCTGGTCTCCAGATGGCAAGAGCTGGCCCTCAGGCTGCAAGAA 611

QY 534 GGGCTCTATAATCATCTGGACCCGGAGACGGGTCAGCAGAAAGGGCGCACCTTGAGTGG 593

Db 612 TGCCAGATTCTCCTCTGGACCCCAAGCACAGGGAAGCAGGTGGCGAGCACCTCGCTGG 671

QY 594 GCACAAGAAACACATCAACTGCCCTCGCTGGGACCGGTATCATCGCGATCCGGAGTGCAG 653

Db 672 CCACAAAGTGGATCACAGGCTGAGCTGGGAGCCCTCCATGTCGAACCTGAGTCCG 731

QY 654 GAACTTGTTCGCCAGTGGAGCGGGGACTGCGCGGATTTGGAGCTGAAAATTTGGGCCA 713

Db 732 CTATGTGCCAGAGCTCCAAGATGGCAGTGTGCGGATCTGGGACACAATGTCAGGCG 791

QY 714 GTGCTTATGAACATTCGCGGACACACAAATGCTGTGACAGAGTGAATGGGTGGAGC 773

Db 792 CTGTGAGCGCATCCTCACCGGGCACACCCAGTCGGTCACTGCTCCTGCGGAGGGGA 851

QY 774 GGGCCTTATTATACATCTCTGGCCCAAGCTCACTGGGTAACAAACATTGCGCTGAGCAC 833

Db 852 CGGGCTTCTCTACTCTGCCCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGAC 911

QY 834 AATCTTGTCCCGAGCTTCTCTGGCCCAAGCTCACTGGGTAACAAACATTGCGCTGAGCAC 893

Db 912 TGTGCTGTGCCGACTCTGCAAGGCCACCGCCACTGGGTGAACACCATGGCCCTCAGCAC 971

QY 894 CGATTACGTCTCGCACCTGGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCC---A 950

Db 972 TGACTATGCCCTGCGCACTGGGGCCTTTGAACCTGCTGAGGCTCAGTTAATCCCCAAGA 1031

QY 951 CCTCAGTTTGACACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACACGCGCGTGTG 1010

Db 1032 CTCCAAGGATCCTTGCAGGAGTTGAAGGAGGGCTCTGAGCCCGATACAACTCGTGGC 1091

QY 1011 CCTGACGAGGTGGAGTCCGTGGTTCCTGTTGGATGACACACACCTCTATCTGTGG-- 1068

Db 1092 GGCCAGGGTCCAGAGAGGCTGGTGTCTGGTCCGACGACTTCACTTATCTGTGGTC 1151

QY 1069 -CGGAACAACCAAGTGGTGGTGGCGGATGACAGGCGCACCAAGACCTGGTCAACGA 1127

Db 1152 CCCAGCAGAGGACAAAAGCCTCTCACTCGGATGACAGGACCAAGCTCTCATCAACCA 1211

QY 1128 TGTGAAATATTCCGCCGGATGTAAGGCTAATTGGCTCTGCTTTCATTGACAAAGTCAGTGG 1187

Db 1212 GGTGCTCTCTCTGACTCCGCGATCGTGGCTAGTGCCTCTTTGACAAAGTCCATCAA 1271  
QY 1188 TCTGTGGCGAGCCAGCGATGGTCAGTACATGGCCACCTTCCGGGGTCAATGTCAGGCTGT 1247  
Db 1272 GCTGTGGGATGGCAGGACGGGGAAGTACCTGGCTCCCTACGCGGCCACGGTGCCT 1331  
QY 1248 TTACACGGTTGCCTGGTCCGGGACTCCCGCTTGATTGTTTCCGGCAGCAAGACTCAAC 1307  
Db 1332 GTACAGATTGCGGTGCTCAGCTGACAGTGGCTCCCTGGTCAGCGGCAGCAGTGACAGCAC 1391  
QY 1308 TCTAAAGTATGGAGTGTGCAGACGAAGAACTGGCCACAGGAGCTGCCTGGACATGCGGA 1367  
Db 1392 ACTGAAGGTGTGGGATGTGAAGGCCAGAGCTGGCCATGGACCTGCCCGGCCACGCGGA 1451  
QY 1368 TGAGGTGTTCCGAGTGGACTGGGCGCCCGATGGCTCTAGAGTTGCCTCTGTTGGCAAGGA 1427  
Db 1452 TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGTGGCAAGTGGTGGGAAGGA 1511  
QY 1428 CAAAGTTATAAGCTATGG 1446  
Db 1512 CAAATGCCTCCGGATATGG 1530

RESULT 14  
BC018399  
LOCUS  
DEFINITION Mus musculus, hypothetical protein similar to beta-transducin family, clone MGC:25690 IMAGE:3491925, mRNA, complete cds.  
ACCESSION BC018399  
VERSION BC018399.1 GI:17390942  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 1784)  
Strausberg,R.  
Direct Submission  
Submitted (03-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-re@mail.nih.gov](mailto:cgapbs-re@mail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 30 Row: b Column: 14.  
Location/Qualifiers  
1. .1784  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="CZECH II"  
/clone="MGC:25690 IMAGE:3491925"  
/tissue\_type="Mammary tumor metastatized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MTV."  
/clone\_lib="NCI\_CGAP\_Lu29"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

CDS  
31. .1488  
/codon\_start=1  
/product="hypothetical protein similar to beta-transducin family"  
/protein\_id="AAH18399.1"  
/db\_xref="GI:17390943"  
/db\_xref="LocusID:54475"  
/translation="MAAAVVEEAAAGDVORILLVQFQDEGGQLLGSPFVDPVDTIPDKL QLVCNALLAQEEPLPLAFYVHDAEIVSSLGKLTLESQSVTEKIVDIYQPQAVFRVRA VTRCTSSLEHSEAVISVAFSPTGKYLASGSDTTFVFWDLSTETPHFTCKGHRHWVL SISWSPDGKKLASGCKNQILLMDPSTGLQVGRVTLTGHSKWITGLSWEPLHNPCECY VASSSKDGSVRVMDTTAGRCERILTGHTQSVTCLRWGGDGLLYSASQDRTIKVWRAHD GVLCRTLQGHGHVNTMALSTDYALRTGAFEPAEATVNAQDLQGSLSKELKERASSRYN LVRGQGPRLVSGSDDTFLFLWSPAEDKKPLARMTGHOALINQVLFSPDSRIVASASF DKSIKLWDGRTGKYLASLRGHVAAVYQIAWSADSRLLVSGSSDSTLKVVDVKAQKLAT DLPGHADDEVYAVDWSPDQQRVASGGKDKCLRWR"

BASE COUNT 397 a 521 c 539 g 327 t  
ORIGIN  
Query Match 30.6%; Score 476.6; DB 10; Length 1784;  
Best Local Similarity 59.2%; Pred. No. 1.9e-108;  
Matches 850; Conservative 0; Mismatches 579; Indels 6; Gaps 2;  
QY 18 GGAGACGGACACGGAGCAAGAGCGCCACGCACATACGATACAGCGCGCCTCGTTTACAC 77  
Db 45 GGTGGAAGAGCGCGCGAGCGGACGTGCAGCGCTGCTCGTACAGTTCCAGGATGAGGG 104  
QY 78 GGGCGAGGAAGCGGCCCGCCCAATCGACTGCCGCGGAGGAATCACTACCCAGCAATTGGG 137  
Db 105 CGGTACGTGCTGGGCTCCCGTTCCGCTGCGAGTCCCCGTGGACATCACGCCGACAGCTGCA 164  
QY 138 ACTGATTGCAACCGCGCTGCTGAAAAACGAGGAAGCCACTCCATATTTGTTTTCGTGGG 197  
Db 165 GCTGGTGTGCAACCGCGCTGCTGGCCCCAGGAGGAGCCCTGCCCTTCTATGTCCA 224  
QY 198 CGAGGATGAGATCAAGAGAGCCCTGGAGGACACACGTTGGACTTGGCGTCAGTGGACACCGA 257  
Db 225 CGATGCTGAGATTGTGTCCTCCCTGGGGAAGACGCTGGAGTCCCAGTCTGTGGAGACAGA 284  
QY 258 AAACGTGATCGATATTGTGTATCAGCCACAGCGCGGTTTCAAAGTGCGCCCGCAGTGACAAG 317  
Db 285 GAAATCGTCGATATCATCTACCAGCCCCCAAGCTGTCTTCAGGGTCCGGGCTGTGACGGG 344  
QY 318 ATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGGTTTCGTGAAATTCAGCCCGCA 377  
Db 345 GTGCACCACTCATTTGGAGGTCACAGCGGAAGCTGTCAATTTCTGTGGCTTCAGCCCCAC 404  
QY 378 TGGTGTCTATCTCGCCAGTGAAGTGGCGGACACACAGTCGAGTGTGGGATCTTAACAC 437  
Db 405 AGGAAAGTACCTGGCCAGTGTTCGGGAGACACCACTGTGCGCTTCTGGGATCTCAGCAC 464  
QY 438 AGAGACACCGCACTTCACCTGCACAGGTCTAAGCAGTGGTGTCTGTGCGTATCTCTGGGC 497  
Db 465 TGAGACACCGCACTTCACATGCAAGGACACCGGCACCTGGGTCTTAGCATATCTCTGGTC 524  
QY 498 TCCGGATGGCAAAACGGTTGGCCAGCGGTTTGCAAGCGGGCTCTATATAATCATCTGGGACCC 557  
Db 525 CCCAGATGGCAAGAACTGGCCTCAGGCTGCAAGAAATGGCCAGATTCTGCTGTGGGACCC 584  
QY 558 GGAGACGGGTGAGCAGAGAAGGGGCGGACCCCTTGAAGTGGGCGACAAAGAACACATCAACTGCCT 617  
Db 585 AAGCACGGGCTGAGGTGGGCGAGGACCCCTCACTGGCCACAGCAAAATGGATCACAGGCCT 644  
QY 618 CGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAACTTGTCTCCGCGAGTGGAGA 677  
Db 645 GAGCTGGGAGCCCCCTTCACATGAACCCCGAGTGCCTACGTGGCCAGCAGTCCCAAAGA 704  
QY 678 CGGGGACTGCCCGGATTTGGGACGTAAAATTGGGCCAGTGCCTTATGAACATTCGCCGGACA 737  
Db 705 CGCAGTGTACGAGTCTGGGATACAAACCGCAGGCCGCTGTGAACGCATCTCTCACGGGACA 764  
QY 738 CACAAATGCTGTGACAGCAGTGAAGATGGGTGGAGCGGCGCTTATTTATACATCTCTCCAA 797



Db 765 CACGCAGTCAGTCACCTGCCTCCGATGGGAGGAGATGGGCTTCTATATATCTGCTTCCCA 824

QY 798 AGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGAATCTTGTCGGGACGTTCTCTGG 857

Db 825 GGACCGAACCATCAAACTCTGGAGGCCCATGACGCTGTGCTGTCGTAATCTCTCAAGG 884

QY 858 CCAAGCTCACTGGTAAACAACATTCGGCTGAGCACCGGATACGTCCTGCGCAGCTGGTCC 917

Db 885 TCATGGCCACTGGTGAACACCATGCGCTTAGCACAGACTATGCCCTGCGCACAGGGGC 944

QY 918 ATTCCATCCGGTGAAGGATCGCTCCAAGAGCC---ACCTCAGTTTGAGCACTAGGAATT 974

Db 945 CTTTGAGCCTGCTGAAGCCACACAGTGAATGCCCAAGACCTGCAGGGTCTTTGAAGGAGCT 1004

QY 975 GCAGGAATCTGCCTGAAGCGCTACACAGCGGTGTCCTTGACGAGGTGGAGTCGCTGGT 1034

Db 1005 GAAGGAGAGGCGATCGAGCCGCTACAACCTCGTGGGGGCCAAGGCCCCAGAGAGGCTGGT 1064

QY 1035 TTCCTGTTCCGATGACAAACACCCCTCTATCTGTGG---CGGAACAACAGAACAAAGTGGT 1091

Db 1065 CTCTGGCTCAGACGACTTCACCTTATTCCTGTGTGTCCTCCAGCAGAGGACAAAGAAGCCCT 1124

QY 1092 TGAGCGCATGACAGGCGACACAGACGTGTCACAGGATGTAATATTCGCCGATGTAAA 1151

Db 1125 CGCCCGGATGACAGGACACACAGGCCCTCATCAACCAAGTCTATCAAGCTGCTCTCTCTGACTCGCG 1184

QY 1152 GCTAATTGGCTCTGCTTCAATTTGACAAAGTCAGTGCCTGTGTGGCGAGCCAGCGATGGTCA 1211

Db 1185 CATCGTGGCCAGTGGCTCCTTCGACAAAGTCTATCAAGCTGTGGGATGGCCGACGGGCAA 1244

QY 1212 GTACATGGCCACCTTCGGGGGTCATGTGCAGGCTGTTTACAGGTTGCCGTGGTCCGCGGA 1271

Db 1245 GTACCTAGCTTCCCTGCGTGGCCACGTTGGCAGCGTGTACCAAGATTGCCCTGAGCTGA 1304

QY 1272 CTCCCGCTTGATTGTTTCCGGCAGCAAGACTCAACTCTAAAGATATGGAGTGTGCAGAC 1331

Db 1305 CAGCGGACTATTACTCAGCGGACGAGTGCAGCACCGTGAAGGTGTGGGATGTGAAGGC 1364

QY 1332 GAAGAAACTGGCACAGGAGCTGCCTGGACATCGCGATGAGTGTTCGGAGTGGACTGGGC 1391

Db 1365 CCAGAAACTGGCCACGGACCTGCCTGGCCATGCCGATGAGTATATGCTGTGACTGGAG 1424

QY 1392 GCCGATGGCTTAGAGTTGCCCTGTTGGTGGCAAGGACAAAGTTATAAGCTATGG 1446

Db 1425 TCCAGATGGCCAGAGAGTAGCAAGCGCGGGAAGACAAAGTGCCTCCGGATATGG 1479

RESULT 15

AX135807

LOCUS

DEFINITION

SEQUENCE 1 from Patent WO0132614.

ACCESSION

AX135807

VERSION

AX135807.1

KEYWORDS

GI:14272042

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1928)

AUTHORS

Utku,N.

TITLE

Novel genes tzap7/a, tzap7/b and tzap7 involved in t cell activation and uses thereof

JOURNAL

Patent: WO 0132614-A 1 10-MAY-2001;

Utku, Nalan (DE)

FEATURES

source

1..1928

/organism="Homo sapiens"

/db\_xref="taxon:9606"

CDS

92..1546

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAC39574.1"

/db\_xref="GI:14272043"

/translation="MVWIADEAVARDVORLLVQFQDEGGQLGSPFDVPVDTTPDRLQ

LVCNALLAQEDPLSLAFFVHDAEIVSSLGKLESAQVETKVLDIILPTQAVFKVRV

TRCTSSLEGHTFAVISVAFSPTGKYLASGSDTTVRFWDLSTETPHFTSGKTHWVLS

IAWSPDGKKLASGCKNSQIFWDPSTGKQIKPLTGHSKWITWLCWEPLHNPESRYL

ASASSGRVDRWDTTAGRCERILTGHQSVTCLRWGGDGLYSASQDRITKVRWHDG

VLCRTLQGHGHVNTMALSTDYALRTGAFEPAEASVNPQDLSQSLQELKERALSRYNL

VRGQGPRLVSGSDDFTLFLWSPAEDKKPLRTMTGHQALINQVLFSFSPSRIVASASFD

KSIKLWDGRTCKYLASLRGHVAAYQIAWSADSRLLVSGSSDSTLKVWDVKAQKLAND

LPGHADEVYAVDWSPDGQVRVSGGKDKCLRIRR"

BASE COUNT 432 a 540 c 550 g 406 t

ORIGIN

Query Match 30.0%; Score 466.2; DB 6; Length 1928;

Best Local Similarity 59.3%; Pred. No. 7.7e-106;

Matches 830; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

QY 54 GATACAGCGCGCCTCGTTTACACGGGCGAGGAAGCGCGCCCGCAATCGACCTGCCGCG 113

Db 139 GCTAGTCAGTTCAGGATGAGGGCGGCGAGCTGCTGGGTTCGCCGTTCGACGTGCCCGT 198

QY 114 AGGAATCACTACCCACCAATTGGGACTGATTGTCACACGCGCTGCTGAAAAACGAGGAAGC 173

Db 199 GGACATACCCCGGACAGGCTGCAGCTCGTTTGCACACGCGCTACTGGCCCCAGGAGATCC 258

QY 174 CACTCCATATTTGTTTTCGTGGGCGAGGATGAGATCAAGAAGACCTGGAGGACACGTT 233

Db 259 CCTGTCACTGGCTTCTTTGTCCACGATGCTGAGATCGTTTCCCTCACTCGGGAAGACGTT 318

QY 234 GGACTTGGCGTCAGTGGACACCCGAAAACGTCATCGATATTGTGTATCAGCCACAGGCGGT 293

Db 319 GGAGTCCCAGGCGAGTGAGACAGAGAAGGTCTTAGACATCATTTTACCCACACAGCTGT 378

QY 294 TTTCAAAGTGCGCCAGTGACAAGATGCACGAGTTCATCGCCGACACCGCGAGGCTGT 353

Db 379 GTTCAAAGTTCGTGCTGTAACACAGATGTACCAGCTCATTTGGAGGACACACCGAGGCTGT 438

QY 354 GGTTCGCTGAATTTTCAGCCCGGATGGTGTCTCATCTCGCCAGTGAAGTGGCGACACCCAC 413

Db 439 TATTTCACTAGCTTCAGCCCACTGGAAAGTATTGGCAAGTGGTCTTGGGGACACTAC 498

QY 414 AGTGGCATTTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGTGCATAAGCA 473

Db 499 AGTCGCTTTGGGATCTCAGCACAGAAAACCTCCACATTTTACATCTAAAGGGCATAACA 558

QY 474 GTGGTTCGTGGCTATCTTGGGTCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAAAGC 533

Db 559 CTGGGTTCTCAGTATGCTTGGTCTCCAGATGGCAAAAACCTTCCTTCAGGATGTAAAAA 618

QY 534 GGGCTCTATAATCATCTGGACCCGGAGACGGTTCAGCAGAGGGCGACCTTGAGTGG 593

Db 619 TAGTCAGATCTTCATTTGGGACCCCAAGCACAGGGAAGCAGATTGGCAAAACCATTAACAGG 678

QY 594 GCACAAGAAACACATCAACTGCCTCGCCTGGGAACCGTATCATTCGGATCCGGAGTGCAG 653

Db 679 GCACTCAAAGTGGATTACATGGGTGTGTGGGAACCTCTCCACCTGAACCCAGAGAGCCG 738

QY 654 GAAACTTGCTTCGCCAGTGGAGACGGGAGCTGCCGGATTTCGGACGTAAATTTGGGCCA 713

Db 739 ATACCTAGCCAGTGCCTCCAGCGGCGCGTCCACCGGATCTGGGACACAACACTGCAGGCCG 798

QY 714 GTGCCTTATGAACATTTGCCGGACACACAATCTGTGTGACAGCAGTGGAGTGGGTGGAGC 773

Db 799 CTGTGAGGGCATCCTCACCGGACACACCCAGTCCGCTCACTGCTCTCCGGTGGGAGGGGA 858

QY 774 GGGCCTTATTTATACATCCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG 833

Db 859 CGGGCTTCTCTACTCTGCCTCCAGGACCCGACCATCAAACTCTGGAGAGCTCATGACGG 918

QY 834 AATCTTGTCCGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATTCGCTGAGCAC 893

Db 919 TGTGCTGTCCGGACTCTGCAAGGCCACCGCCACTGGGTGAACACCATGGCCCTCAGCAC 978

QY 894 CGATTACGTCCTGCGCAGTGTGCTCCATTCCATCCGGTGAAGGATCGCTCCAAAGAGCC---A 950



[illegible]

Search completed: June 2, 2003, 04:13:27  
Job time : 3999 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 2, 2003, 01:13:14 ; Search time 378 Seconds  
(without alignments)  
9264.173 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattccccaaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002:\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1555	100.0	1555	21	AAA27739
2	1487.8	95.7	1992	23	ABL05339
3	1285.8	82.7	4164	23	ABL05338
4	601.2	38.7	6223	23	ABL05232
5	588.2	37.8	4141	23	ABL05214
6	478	30.7	1859	22	AAF85075
7	477.4	30.7	1853	22	AAH14166
8	477.4	30.7	1854	22	AAF58316
9	477.4	30.7	2638	24	ABN59719
					Drosophila Notchle
					Drosophila melanog
					Drosophila melanog
					Drosophila melanog
					Drosophila melanog
					Nucleotide sequenc
					Human cDNA sequenc
					Human GTP-binding
					Novel human coding

10	466.2	30.0	1928	22	AAF85073	Nucleotide sequenc
11	333.4	21.4	1726	21	AAC48725	Arabidopsis thalia
12	331.8	21.3	1643	21	AAC39906	Arabidopsis thalia
13	249.4	16.0	1170	22	AAF85074	Nucleotide sequenc
14	176	11.3	399	21	AAH30571	Human colon cancer
15	164	10.5	547	22	AAH05727	Human cDNA clone (
16	155.8	10.0	1722	20	AAH05727	DNA encoding a hum
17	153.6	9.9	573	22	AAH98017	Murine 7-transmemb
18	123.8	8.0	540	21	AAF07777	Fusarium venenatum
19	120.4	7.7	378	22	AAF67530	Novel human polynu
20	80.4	5.2	316	21	AAF11429	Aspergillus niger
21	79.4	5.1	2984	22	AAK94739	Human full-length
22	79.4	5.1	3025	22	ABA06363	Human cDNA SEQ ID
23	78.4	5.0	300	20	AAZ14360	Human gene express
24	78.4	5.0	764	22	AAK91829	Human cDNA 5'-end
25	78.4	5.0	764	22	AAK93229	Human cDNA clone r
26	78.4	5.0	1950	22	AAK52860	Human polynucleoti
27	69	4.4	1925	22	AAK51876	Human polynucleoti
28	68.4	4.4	279	23	ABV55135	Human prostate exp
29	68.4	4.4	2977	24	ABQ72680	Human MDPF encodin
30	68.2	4.4	1464	24	ABN59758	Novel human coding
31	68.2	4.4	1751	22	AAF58329	Human GTP-binding
32	68.2	4.4	2064	21	AAF22375	Human secreted pro
33	64	4.1	1591	23	ABL23519	Drosophila melanog
34	62.6	4.0	462	23	ABV57002	Human prostate exp
35	62.2	4.0	559	23	ABV54541	Human prostate exp
36	62	4.0	1810	21	AAC77522	Human ORFX ORF3077
37	61.8	4.0	2359	15	AAQ70726	TATA-binding prote
38	61.4	3.9	821	22	AAK92334	Human cDNA 5'-end
39	61.4	3.9	821	22	AAK93787	Human cDNA clone r
40	61.4	3.9	1890	22	AAK94610	Human full-length
41	60.2	3.9	2359	17	AAT42212	Drosophila TATA-bi
42	60.2	3.9	2359	18	AAT79603	TATA-binding prote
43	60.2	3.9	2826	23	ABL13261	Drosophila melanog
44	60.2	3.9	3965	23	ABL11658	Drosophila melanog
45	60.2	3.9	4826	23	ABL13260	Drosophila melanog

ALIGNMENTS

RESULT 1

AAA27739

ID AAA27739 standard; cDNA; 1555 BP.

XX AAA27739;

AC AAA27739;

XX 29-AUG-2000 (first entry)

XX Drosophila Notchless (Nle) cDNA.

DE Notchless; Nle gene; Notch; signalling; neurodegenerative disease;

XX cancer; diagnosis; cytostatic; neuroprotective; gene therapy; ss.

KW Drosophila melanogaster.

XX Drosophila melanogaster.

OS Drosophila melanogaster.

XX Key

XX Location/Qualifiers

FT CDS

FT 13..1455

FT /\*tag= a

FT /transl\_except= (pos:859..861, aa:His)

XX WO200026364-A1.

XX PD

XX 11-MAY-2000.

XX 03-NOV-1999; 99WO-IB01891.

XX 03-NOV-1998; 98GB-0024045.

XX (EUMO-) EURO MOLECULAR BIOLOGY LAB.

XX Cohen S, Bouwmeester A, Royet J;

XX

DR WPI; 2000-365613/31.  
DR P-PSDB; AAY79678.

PT: Novel Notchless protein and nucleic acids encoding them useful for  
PT: treating and preventing cancer and neurodegenerative diseases -  
XX  
PS Claim 19; Page 44-45; 52pp; English.

The present sequence is that of cDNA encoding Notchless, a novel protein of *Drosophila*. The cDNA was isolated from an eye disc cDNA library by screening with a genomic DNA fragment that flanks the p element insertion site in *Drosophila* 1(2)AAK13714. Notchless was identified in a screen for dominant modifiers of a Notch mutant phenotype in the *Drosophila* wing. The mutant dominantly suppressed the wing notching phenotype of notchoid mutations, and the Notchless protein was shown to bind to the cytoplasmic domain of Notch. Notchless modified Notch signalling activity in a variety of Notch-dependent signalling process in both *Drosophila* and *Xenopus* embryos. The Notchless protein, and nucleic acids encoding it, can be used in methods for the diagnosis and therapy of certain diseases, particularly cancer and neurodegenerative diseases (claimed). A probe capable of screening for the Notchless gene (Nle), a cloning or expression vector comprising Notchless DNA, cDNA or RNA, a host cell, a transgenic animal, and a method for production of Notchless protein in the host cell are also claimed.

SQ Sequence 1555 BP; 418 A; 373 C; 448 G; 316 T; 0 other;

Query Match	100.0%;	Score 1555;	DB 21;	Length 1555;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1555;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

RESULT 2  
ABL05339

ID	ABL05339	standard;	CDNA;	1992	BP.
1	1	1	1	1	1

AC ABL05339;

XX

DT 26-MAR-2002 (first entry)

XX

22



XX Drosophila; developmental biology; cell signalling; insecticide; 424 TGGGATCTTAACACAGACAGACCCGACCTTCACCTGCACAGGTCATAAGCAGTGGTTCG 483  
KW pharmaceutical; gene; ss. 439 TGGGATCTTAACACAGACAGACCCGACCTTCACCTGCACAGGTCATAAGCAGTGGTTCG 498  
XX Drosophila melanogaster. 484 TGGGATCTTCCGCTCCGGATGCAACACGGTTGGCCAGCGGTTGCAAAAGCGGCTCTATA 543  
OS WO200171042-A2. 499 TGGGATCTTCCGCTCCGGATGCAACACGGTTGGCCAGCGGTTGCAAAAGCGGCTCTATA 558  
XX 27-SEP-2001. 544 ATCATCTGGACCCGGAGACGGGTCAGCAGAGGGGGCGACCTTGAGTGGGCACAAGAAA 603  
XX 23-MAR-2001; 2001WO-US09231. 559 ATCATCTGGACCCGGAGACGGGTCAGCAGAGGGGGCGACCTTGAGTGGGCACAAGAAA 618  
XX 23-MAR-2000; 2000US-191637P. 604 CACATCAACTGCCTCGCTGGGAAACCGTATCATCGCGATCCGGAGTGCAGGAACCTTGCT 663  
PR 11-JUL-2000; 2000US-0614150. 619 CACATCAACTGCCTCGCTGGGAAACCGTATCATCGCGATCCGGAGTGCAGGAACCTTGCT 678  
XX (PEKE ) PE CORP NY. 664 TCCGCCAGTGGAGACGGGGACTGCGCGATTGGGACGTAATAATTGGGGCCAGTGCCTTATG 723  
XX Venter JC, Adams M, Li PWD, Myers EW; 679 TCCGCCAGTGGAGACGGGGACTGCGCGATTGGGACGTAATAATTGGGGCCAGTGCCTTATG 738  
PI P-PSDB; ABB61236. 724 AACATGCCGACACACAAATGCTGTGACAGCAGTGGGTTGGAGCGGCGCTTATT 783  
XX New isolated nucleic acid detection reagent for detecting 1000 or more 739 AACATGCCGACACACAAATGCTGTGACAGCAGTGGGTTGGAGCGGCGCTTATT 798  
PT genes from Drosophila and for elucidating cell signalling and cell-cell 784 TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGAATCTTGTGC 843  
PT interactions - 799 TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGAATCTTGTGC 858  
XX Claim 1; SEQ ID NO 10499; 2lpp + Sequence Listing; English. 844 CGGACGTTCTTGGCCAAAGCTCAGTGGGTAACAAACATTCGCTGAGCACCGGATACGTC 903  
XX The invention relates to an isolated nucleic acid detection reagent 859 CGGACGTTCTTGGCCAGCTCAGTGGGTAACAAACATTCGCTGAGCACCGGATACGTC 918  
CC capable of detecting 1000 or more genes from Drosophila. The invention is 904 CTGCGCACTGGTCCATTCCATCCGTTGAAGGATCGCTCCAAAGAGCCACCTCAGTTTGAGC 963  
CC useful in developmental biology and in elucidating cell signalling and 919 CTGCGCACTGGTCCATTCCATCCGTTGAAGGATCGCTCCAAAGAGCCACCTCAGTTTGAGC 978  
CC cell-cell interactions in higher eukaryotes for the development of 964 ACTGAGGAATTGCAGGAATCTGCCCTTGAAGCGCTACAGGCGCTGTCCTGACGAGGTG 1023  
CC insecticides, therapeutics and pharmaceutical drugs. The invention 979 ACTGAGGAATTGCAGGAATCTGCCCTTGAAGCGCTACAGGCGCTGTCCTGACGAGGTG 1038  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA 1024 GAGTCGCTGGTTTCCCTGTTGGATGACAAACACCCCTCATCTGTGGCGGAACACCAAGAAC 1083  
CC sequences (ABL01840-ABL16175) and the encoded proteins 1039 GAGTCGCTGGTTTCCCTGTTGGATGACAAACACCCCTCATCTGTGGCGGAACACCAAGAAC 1098  
CC (ABB57737-ABB72072). 1084 AAGTGCCTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGATGTGAATATTCGCCG 1143  
CC The sequence data for this patent did not form part of the printed 1099 AAGTGCCTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGATGTGAATATTCGCCG 1158  
CC specification, but was obtained in electronic format directly from WIPO 1144 GATGTAAAGCTAATGCGTCTGCTTCAATTGACAAGTCAAGTCAAGTGGCGGAGCCAGC 1203  
CC at ftp.wipo.int/pub/published\_pct\_sequences. 1159 GATGTAAAGCTAATGCGTCTGCTTCAATTGACAAGTCAAGTCAAGTGGCGGAGCCAGC 1218  
XX Sequence 1992 BP; 482 A; 492 C; 585 G; 433 T; 0 other; 1204 GATGTCAGTACATGGCCACCTTCCGGGGTCATGTGACGCTGTTTACACGGTTGCCCTGG 1263  
SQ Query Match 95.7%; Score 1487.8; DB 23; Length 1992; 1219 GATGTCAGTACATGGCCACCTTCCGGGGTCATGTGACGCTGTTTACACGGTTGCCCTGG 1278  
Best Local Similarity 99.3%; Pred. No. 0; 1264 TCCGCGGACTCCCGCTTGATGTTTCCGGGACCAAGACTCAACTCTAAAAGTATGGAGT 1323  
Matches 1505; Conservative 0; Mismatches 7; Indels 3; Gaps 1; 1279 TCCGCGGACTCCCGCTTGATGTTTCCGGGACCAAGACTCAACTCTAAAAGTATGGAGT 1338  
QY 7 CAAAAAATGCAGGACGCGACACGGAGCAAGAGCGCACGCCACATACGATACAGGCGCGC 66 1324 GTGCAGACGAAGAAACTGGCACAGGAGCTGCCTGGACATCGCGGATGAGGTGTCGGAGTG 1383  
Db 19 CAGAAAATGCAGGACGCGACACGGAGCAAGAGCGCACGCCACATACGATACAGGCGCGC 78 1339 GTGCAGACGAAGAAACTGGCACAGGAGCTGCCTGGACATCGCGGATGAGGTGTCGGAGTG 1398  
QY 67 CTCGTTT---ACACGGGCGAGGAAGCGGCGCCGCAATCGACCTGCCGCGAGGAATCACT 123 1384 GACTGGGCGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAGTTATAAAGCTA 1443  
Db 79 CTCGTTTTCGGACACGCGGCGAGGAAGCGGCGCCGCAATCGACCTGCCGCGAGGAATCACT 138 1399 GACTGGGCGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAGTTATAAAGCTA 1458  
QY 124 ACCCAGCAATTGGGACTGATTGCAACGCGCTGTGAAAAACGAGGAAGCCACTCCATAT 183 1444 TGGGCTTATTAAACAAATCATTAACCTTGACACGGTGAAGAAATACTTAGGAATAAAGTAA 1503  
Db 139 ACCCAGCAATTGGGCTGTGATTGCAACGCGCTGTGAAAAACGAGGAAGCCACTCCATAT 198 1459 TGGGCTTATTAAACAAATCATTAACCTTGACACGGTGAAGAAATACTTAGGAATAAAGTAA 1518  
QY 184 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 243 304 CGCCCAAGTGAAGATGCACGAGTTCCATGCCGGACACGCCGAGGCTGTGGTTTCGCTG 363  
Db 199 TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 258 319 CGCCCAAGTGAAGATGCACGAGTTCCATGCCGGACACGCCGAGGCTGTGGTTTCGCTG 378  
QY 244 TCAGTGGACACCGAAACGCTGATCGATATTGTGTATCAGCCACAGCGGTTTTCRAAGTG 303 364 AATTTCAGCCCGGATGGTCTCATCTGCCAGTGGAGTGGGACACACAGTGCATTTG 423  
Db 259 TCAGTGGACACCGAAACGCTGATCGATATTGTGTATCAGCCACAGCGGTTTTCRAAGTG 318 379 AATTTCAGCCCGGATGGTCTCATCTGCCAGTGGAGTGGGACACACACAGTGCATTTG 438  
QY 304 CGCCCAAGTGAAGATGCACGAGTTCCATGCCGGACACGCCGAGGCTGTGGTTTCGCTG 363  
Db 319 CGCCCAAGTGAAGATGCACGAGTTCCATGCCGGACACGCCGAGGCTGTGGTTTCGCTG 378  
QY 364 AATTTCAGCCCGGATGGTCTCATCTGCCAGTGGAGTGGGACACACAGTGCATTTG 423  
Db 379 AATTTCAGCCCGGATGGTCTCATCTGCCAGTGGAGTGGGACACACACAGTGCATTTG 438

QY	424	TGGGATCTTAAACACAGACAGACCCGACCTTCACCTGCACAGGTCATAAGCAGTGGTTCG	483
Db	439	TGGGATCTTAAACACAGACAGACCCGACCTTCACCTGCACAGGTCATAAGCAGTGGTTCG	498
QY	484	TGGGATCTTCCGCTCCGGATGCAACACGGTTGGCCAGCGGTTGCAAAAGCGGCTCTATA	543
Db	499	TGGGATCTTCCGCTCCGGATGCAACACGGTTGGCCAGCGGTTGCAAAAGCGGCTCTATA	558
QY	544	ATCATCTGGACCCGGAGACGGGTCAGCAGAGGGGGCGACCTTGAGTGGGCACAAGAAA	603
Db	559	ATCATCTGGACCCGGAGACGGGTCAGCAGAGGGGGCGACCTTGAGTGGGCACAAGAAA	618
QY	604	CACATCAACTGCCTCGCTGGGAAACCGTATCATCGCGATCCGGAGTGCAGGAACCTTGCT	663
Db	619	CACATCAACTGCCTCGCTGGGAAACCGTATCATCGCGATCCGGAGTGCAGGAACCTTGCT	678
QY	664	TCCGCCAGTGGAGACGGGGACTGCGCGATTGGGACGTAATAATTGGGGCCAGTGCCTTATG	723
Db	679	TCCGCCAGTGGAGACGGGGACTGCGCGATTGGGACGTAATAATTGGGGCCAGTGCCTTATG	738
QY	724	AACATGCCGACACACAAATGCTGTGACAGCAGTGGGTTGGAGCGGCGCTTATT	783
Db	739	AACATGCCGACACACAAATGCTGTGACAGCAGTGGGTTGGAGCGGCGCTTATT	798
QY	784	TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGAATCTTGTGC	843
Db	799	TATACATCTCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGAATCTTGTGC	858
QY	844	CGGACGTTCTTGGCCAAAGCTCAGTGGGTAACAAACATTCGCTGAGCACCGGATACGTC	903
Db	859	CGGACGTTCTTGGCCAGCTCAGTGGGTAACAAACATTCGCTGAGCACCGGATACGTC	918
QY	904	CTGCGCACTGGTCCATTCCATCCGTTGAAGGATCGCTCCAAAGAGCCACCTCAGTTTGAGC	963
Db	919	CTGCGCACTGGTCCATTCCATCCGTTGAAGGATCGCTCCAAAGAGCCACCTCAGTTTGAGC	978
QY	964	ACTGAGGAATTGCAGGAATCTGCCCTTGAAGCGCTACAGGCGCTGTCCTGACGAGGTG	1023
Db	979	ACTGAGGAATTGCAGGAATCTGCCCTTGAAGCGCTACAGGCGCTGTCCTGACGAGGTG	1038
QY	1024	GAGTCGCTGGTTTCCCTGTTGGATGACAAACACCCCTCATCTGTGGCGGAACACCAAGAAC	1083
Db	1039	GAGTCGCTGGTTTCCCTGTTGGATGACAAACACCCCTCATCTGTGGCGGAACACCAAGAAC	1098
QY	1084	AAGTGCCTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGATGTGAATATTCGCCG	1143
Db	1099	AAGTGCCTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGATGTGAATATTCGCCG	1158
QY	1144	GATGTAAAGCTAATGCGTCTGCTTCAATTGACAAGTCAAGTCAAGTGGCGGAGCCAGC	1203
Db	1159	GATGTAAAGCTAATGCGTCTGCTTCAATTGACAAGTCAAGTCAAGTGGCGGAGCCAGC	1218
QY	1204	GATGTCAGTACATGGCCACCTTCCGGGGTCATGTGACGCTGTTTACACGGTTGCCCTGG	1263
Db	1219	GATGTCAGTACATGGCCACCTTCCGGGGTCATGTGACGCTGTTTACACGGTTGCCCTGG	1278
QY	1264	TCCGCGGACTCCCGCTTGATGTTTCCGGGACCAAGACTCAACTCTAAAAGTATGGAGT	1323
Db	1279	TCCGCGGACTCCCGCTTGATGTTTCCGGGACCAAGACTCAACTCTAAAAGTATGGAGT	1338
QY	1324	GTGCAGACGAAGAAACTGGCACAGGAGCTGCCTGGACATCGCGGATGAGGTGTCGGAGTG	1383
Db	1339	GTGCAGACGAAGAAACTGGCACAGGAGCTGCCTGGACATCGCGGATGAGGTGTCGGAGTG	1398
QY	1384	GACTGGGCGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAGTTATAAAGCTA	1443
Db	1399	GACTGGGCGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAGTTATAAAGCTA	1458
QY	1444	TGGGCTTATTAAACAAATCATTAACCTTGACACGGTGAAGAAATACTTAGGAATAAAGTAA	1503
Db	1459	TGGGCTTATTAAACAAATCATTAACCTTGACACGGTGAAGAAATACTTAGGAATAAAGTAA	1518

```
QY 1504 AACGTCCTGAGTAAA 1518
Db 1519 AACGTCCTGAGTAAA 1533

RESULT 3
ABL05338
ID ABL05338 standard; cDNA; 4164 BP.
XX
AC ABL05338;
XX
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 10496.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB61235.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 10496; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4164 BP; 947 A; 1030 C; 1101 G; 1086 T; 0 other;

Query Match 82.7%; Score 1285.8; DB 23; Length 4164;
Best Local Similarity 89.2%; Pred. No. 1.7e-279;
Matches 1505; Conservative 0; Mismatches 7; Indels 175; Gaps 4;

QY 7 CAAAAAATGCAGGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGGC 66
Db 1019 CAGAAATGCAGGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGGC 1078

QY 67 CTCGTTT---ACACGGCGGAGGAGCCGGCCGCCCAATCGACCTGCCGGCAGGAATCACT 123
Db 1079 CTCGTTTCGGACACGGCGGAGGAGCGGGCCGCCCAATCGACCTGCCGGCAGGAATCACT 1138

QY 124 ACCCAGCAATTGGGACTGATTTGCAACGGCTGCTGAAAACGAGGAGGAGCCACTCCCATAT 183
Db 1139 ACCCAGCAATTGGGCTGATTTGCAACGGCTGCTGAAAACGAGGAGGAGCCACTCCCATAT 1198

QY 184 TTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 243
```

```
Db 1199 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAAGAGCCTGGAGGACACGTTGGACTTGGCG 1258
QY 244 TCAGTGGACACCGAAAAACGTGATCGATATGTGTATCAGCCACAGCGGTTTTTCAAAGTG 303
Db 1259 TCAGTGGACACCGAAAAACGTGATCGATATGTGTATCAGCCACAGCGGTTTTTCAAAGTG 1318
QY 304 CGCCACAGTGACAAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 363
Db 1319 CGCCACAGTGACAAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 1378
QY 364 AATTTCAGCCCGGATGGTCTCATCTCGCCAGTGGAGTGGCGACACACAGTGGATTG 423
Db 1379 AATTTCAGCCCGGATGGTCTCATCTCGCCAGTGGAGTGGCGACACACAGTGGATTG 1438
QY 424 TGGGATCTTAAACACAGACACACCGCACCTTCACCTGCACAGGTCTAAGCAGTGGGTTCTG 483
Db 1439 TGGGATCTTAAACACAGACACACCGCACCTTCACCTGCACAGGTCTAAGCAGTGGGTTCTG 1498
QY 484 TCGGTATCCTTGGGCTCCGGATGGCAACCGTTGGCCACGCGTTGCAAGCGGGCTCTATA 543
Db 1499 TCGGTATCCTTGGGCTCCGGATGGCAACCGTTGGCCACGCGTTGCAAGCGGGCTCTATA 1558
QY 544 ATCATCTGGGACCCCGGAGACGGGTGAGCAAGAGGGCGACCCCTTGAGTGGGCACAAAGAAA 603
Db 1559 ATCATCTGGGACCCCGGAGACGGGTGAGCAAGAGGGCGACCCCTTGAGTGGGCACAAAGAAA 1618
QY 604 CACATCAACTGCTCGCTCGCTGGGAAACCGTATCATCGCATCCGGAGTCCAGGAAACCTGCT 663
Db 1619 CACATCAACTGCTCGCTCGCTGGGAAACCGTATCATCGCATCCGGAGTCCAGGAAACCTGCT 1678
QY 664 TCCGCCAGTGGAGACGGGACTCCGGATTGGGACGTGAAATTTGGCCAGTGCCTTATG 723
Db 1679 TCCGCCAGTGGAGACGGGACTCCGGATTGGGACGTGAAATTTGGCCAGTGCCTTATG 1738
QY 724 AACATTGCCCGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGCGGGCCTTATT 783
Db 1739 AACATTGCCCGACACACAAATGCTGTGACAGCAGTGAGATGGGTGGAGCGGGCCTTATT 1798
QY 784 TATACATCCTCCAAAGATCGACAGTGAAAGATGTGGCGAGCAGCTGATGGAATCTTGTC 843
Db 1799 TATACATCCTCCAAAGATCGACAGTGAAAGATGTGGCGAGCAGCTGATGGAATCTTGTC 1858
QY 844 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATTTGCGTGAACCCGATFACGTC 903
Db 1859 CGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATTTGCGTGAACCCGATFACGTC 1918
QY 904 CTGCGCACTGCTCCATTCATCCGGTGAAGGATCGCTCCAAAGAGCCACCTCAGTTGAGC 963
Db 1919 CTGCGCACTGCTCCATTCATCCGGTGAAGGATCGCTCCAAAGAGCCACCTCAGTTGAGC 1978
QY 964 A-----C 965
Db 1979 AGTAAGAAATTCATATGCTCTATATATATGCTGTATATCCTAAAAGACATGTTTACAGC 2038
QY 966 TGAGGAATTCAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTGCCCTGACGAGGTGGA 1025
Db 2039 TGAGGAATTCAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTGCCCTGACGAGGTGGA 2098
QY 1026 GTCGCTGGTTTCCTGTTCCGATGACAAACACCCCTCTATCTGTGGCGGAACAACCAAGACAA 1085
Db 2099 GTCGCTGGTTTCCTGTTCCGATGACAAACACCCCTCTATCTGTGGCGGAACAACCAAGACAA 2158
QY 1086 GTGCGTTGAGCGCATGACAGGGCACCAAGAACGTGGTCAACGATGTGAAATATTCGCCGGA 1145
Db 2159 GTGCGTTGAGCGCATGACAGGGCACCAAGAACGTGGTCAACGATGTGAAATATTCGCCGGA 2218
QY 1146 TGTAAAGCTAATTCGCTGCTTCAATTGACAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 1205
Db 2219 TGTAAAGCTAATTCGCTGCTTCAATTGACAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 2278
QY 1206 TGGTCAGTACATGGCCACCTCCGGGGTCAATGTCAGGCTGTTTACAGGTTGCTGCTGCTC 1265
```

Db 2279 TGGTCAGTACATGGCCACCTTCCGGGGTCATGTCAGGCTGTTTACACGGTGCCTGGTC 2338  
QY 1266 CGCGGACTCCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAA----- 1313  
Db 2339 CGCGGACTCCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAAAGGTTAGCACGT 2398  
QY 1314 -----AGTATGGAGTGTGCAGA 1330  
Db 2399 AGTTTCAGTTTTCTACGCTAGCTAATAACATATCCTCTTGCTTAGTATGGAGTGTGCAGA 2458  
QY 1331 CGAAGAAACTGGCACAGGAGCTGCCTGGACATGCGGATGAGCTGTTCCGAGTGGACTGGG 1390  
Db 2459 CGAAGAAACTGGCACAGGAGCTGCCTGGACATGCGGATGAGCTGTTCCGAGTGGACTGGG 2518  
QY 1391 CGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAGCTTATAAAGCT----- 1442  
Db 2519 CGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAGCTTATAAAGCTTAAGTTT 2578  
QY 1443 -----ATGGGCTTAA 1451  
Db 2579 GAGTAACGAGGTTTCTATATAACAATATACAGACTCTAATGCTATCTTTTCAGATGGGCTTA 2638  
QY 1452 TTAACAAATCATTAACCTGTACACGGTAAGAAATACTTTAGGAATAAAGTAAACGTCCT 1511  
Db 2639 TTAACAAATCATTAACCTGTACACGGTAAGAAATACTTTAGGAATAAAGTAAACGTCCT 2698  
QY 1512 GAGTAAA 1518  
Db 2699 GAGTAAA 2705

RESULT 4

ID ABL05232 standard; cDNA; 6223 BP.

XX ABL05232;

AC ABL05232;

XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster expressed polynucleotide SEQ ID NO 10178.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

PF 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

XX P-PSDB; ABB61129.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

PS Claim 1; SEQ ID NO 10178; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 6223 BP; 1458 A; 1588 C; 1574 G; 1603 T; 0 other;

Query Match 38.7%; Score 601.2; DB 23; Length 6223;  
Best Local Similarity 99.0%; Pred. No. 1.8e-125;  
Matches 616; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

QY 7 CAAAAAATGCAGAGACGGACACGGAGCAAGAGCCACGCCACATACATACAGGCGCGC 66  
Db 5602 CAGAAAATGCAGAGACGGACACGGAGCAAGAGCCACGCCACATACATACAGGCGCGC 5661  
QY 67 CTCGTTT---ACACGGGGGAGGAAGCCGCCGCCCAATGACCTGCCGGCAGGAATCACT 123  
Db 5662 CTCGTTTCCGACACGGGGGAGGAAGCGGGCCGCCCAATGACCTGCCGGCAGGAATCACT 5721  
QY 124 ACCCAGCAATTTGGGACTGATTTGCAACGCGCTGCTGAAACAGGAGGAGCCACTCCATAT 183  
Db 5722 ACCCAGCAATTTGGTCTGATTTGCAACGCGCTGCTGAAACAGGAGGAGCCACTCCATAT 5781  
QY 184 TTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCTGGAGGACACGCTTGGACTTGGCG 243  
Db 5782 TTGTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCTGGAGGACACGCTTGGACTTGGCG 5841  
QY 244 TCAGTGGACACCGAAAAACGTGATCGATATGTGTATCAGCCACACAGGCGGTTTCAAAGTG 303  
Db 5842 TCAGTGGACACCGAAAAACGTGATCGATATGTGTATCAGCCACACAGGCGGTTTCAAAGTG 5901  
QY 304 CGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCAGGCTGTGGTTTCGCTG 363  
Db 5902 CGCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCAGGCTGTGGTTTCGCTG 5961  
QY 364 AATTTCAGCCCGGATGGTGTCTCATCTCGCCAGTGGGAAGTGGCGACACACAGTGGATTG 423  
Db 5962 AATTTCAGCCCGGATGGTGTCTCATCTCGCCAGTGGGAAGTGGCGACACACAGTGGATTG 6021  
QY 424 TGGGATCTTAACACAGACACCGGCTTACCTGCACAGGTCATAGCAGTGGTCTCTG 483  
Db 6022 TGGGATCTTAACACAGACACCGGCTTACCTGCACAGGTCATAGCAGTGGTCTCTG 6081  
QY 484 TGCGTATCTGGCTCCGGATGGCAACCGTTGGCCAGCGGTTGCAAGCGGGCTCTATA 543  
Db 6082 TGCGTATCTGGCTCCGGATGGCAACCGTTGGCCAGCGGTTGCAAGCGGGCTCTATA 6141  
QY 544 ATCATCTGGACCCGGAGACGGGTTCAGCAGAGGGGGCGGAGCCCTTGAGTGGGCACAGAAA 603  
Db 6142 ATCATCTGGACCCGGAGACGGGTTCAGCAGAGGGGGCGGAGCCCTTGAGTGGGCACAGAAA 6201  
QY 604 CACATCAACTGCCTCGCCTGGG 625  
Db 6202 CACATCAACTGCCTCGCCTGGG 6223

RESULT 5

ID ABL05214 standard; cDNA; 4141 BP.

XX ABL05214;

AC ABL05214;

XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster expressed polynucleotide SEQ ID NO 10124.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

OS

XX













Db	1329	ACTGAAGGTGTGGGATGTGAAGGCCAGAGCTGGCCATGGACCTGCCCGGCCACGGCGA	1388
QY	1368	TGAGGTGTTTCGGAGTGGACTGGGGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGA	1427
Db	1389	TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA	1448
QY	1428	CAAAGTTATAAAGCTATGG	1446
Db	1449	CAAATGCCCTCCGGATATGG	1467
RESULT 9			
ABN59719	ID	ABN59719 standard; cDNA; 2638 BP.	
XX	AC	ABN59719;	
XX	DT	28-JUN-2002 (first entry)	
XX	DE	Novel human coding sequence SEQ ID NO: 130.	
XX	KW	Human; antianaemic; vulnery; antiinflammatory; immunomodulator;	
KW	KW	antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;	
KW	KW	neuroprotective; antiparkinsonian; protein therapy; EST;	
XX	XX	expressed sequence tag; gene; ss.	
OS	XX	Homo sapiens.	
XX	PN	WO200222660-A2.	
XX	PD	21-MAR-2002.	
XX	PF	10-SEP-2001; 2001WO-US26015.	
XX	PR	11-SEP-2000; 2000US-0659671.	
XX	PA	(HYSE-) HYSEQ INC.	
PI	PI	Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;	
PI	PI	Xue AJ, Yang Y, Wehrman T, Drmanac RT;	
XX	DR	WPI; 2002-292408/33.	
XX	PT	P-PSDB; ABB97306.	
PT	PT	An isolated polynucleotide for treating diseases associated with its	
XX	PS	encoded polypeptide such as cancer and multiple sclerosis -	
XX	XX	Claim 1; SEQ ID NO 130; 509pp; English.	
CC	CC	The present invention provides the protein and coding sequences of 444	
CC	CC	novel human proteins. These were isolated from expressed sequences tags	
CC	CC	(ESTs). They can be used to stimulate cell growth, to regulate	
CC	CC	haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth	
CC	CC	e.g. in burn treatment, to regulate the immune system e.g. to treat	
CC	CC	multiple sclerosis, to regulate activin or inhibit e.g. to treat	
CC	CC	infertility, to regulate haemostasis or thrombolysis e.g. to treat	
CC	CC	stroke and cancer, to screen for drugs, to treat inflammatory conditions	
CC	CC	e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.	
CC	CC	Parkinson's disease. The present sequence is a coding sequence of the	
XX	SQ	invention.	
Sequence 2638 BP; 566 A; 767 C; 764 G; 541 T; 0 other;			
Query Match			
Best Local Similarity 30.7%; Score 477.4; DB 24; Length 2638;			
Matches 837; Conservative 0; Mismatches 556; Indels 6; Gaps 2;			
QY	54	GATACAGGGCGCCTCGTTTACACGGGGGAGGAAGCCGGCCGCCAATCGACCTGCCGGC	113
Db	132	GCTAGTGCAGTTCAGGATGAGGGCGGGACGCTGCTGGGTCCCCGTTCCGACGTGCCCGT	191
QY	114	AGGATCACTACCCAGCAATTGGGACTGATTGCAACGGCGCTGCTGAAAAACGAGGAAGC	173

Db	192	GGACATCACCCCGGACAGGCTGCAGCTCGTGTGCAACGGCGTACTGGCCAGGAGATCC	251
QY	174	CACCTCCATATTGTTTTCGTGGCGAGGATGAGATCAAGAAGAGAGCTGGAGGACACGTT	233
Db	252	CCTGCCACTGGCTTCTTTGTCCACGATGCTGAGATCGTCTCCTCACTGGGGAAGACGTT	311
QY	234	GGACTTGGCGTCAGTGGACACCCGAAACGTTGATCGATATTGTGTATCAGCCACAGCGGT	293
Db	312	GGAGTCCCAGGCAGTGGACAGAGAGGTCTTAGACATCATCTACAGCCACAGGCTAT	371
QY	294	TTTCAAAGTGGCCCGCAGTGCACAAAGATGCACGAGTTCCATGCCGGGACACGCCAGGCTGT	353
Db	372	CTTCAGAGTCCGGGCTGTGACTCGCTGCACCAAGTACCTGGCCAGTGGCTCTGGAGGCAGT	431
QY	354	GGTTTCGCTGAATTTCAGCCCGGATGGTGTCTCATCTGCCAGTGGGAAGTGGGACACAC	413
Db	432	CATTTCTGTGGCCTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACAC	491
QY	414	AGTGGATTGTGGGATCTTAACACAGACACACCGGACTTCACCTGCACAGGTCATAAGCA	473
Db	492	CGTGGCTTCTGGGATCTCAGACACAGACACACCAATTCACATGCAAGGACACAGACA	551
QY	474	GTGGGTTCTGTGCGTATCCTGGGCTCCGGATGGCAACGGTTGCCAGCGGTTGCCAAGC	533
Db	552	CTGGGTCTTAGTATATCTCTCCAGATGGCAAGAGTGGCCCTCAGGCTGCAAGAA	611
QY	534	GGGCTCTATAATCATCTGGGACCCGGAGACGGGTGTCAGAGAAGGGGACCCCTGAGTGG	593
Db	612	TGGCCAGATTCTCTCTGGGACCCCAAGCACAGGAGGAGTGGGAGGACCCCTCGCTGG	671
QY	594	GCACAAGAAACACATCAACTGCCCTCGCTGGGAACCGATATCATCGGATCCGGAGTGCAG	653
Db	672	CCACAGCAAGTGGATCACAGGCTGAGCTGGAGCCCTCCATGCGAAACCTGAGTGCCG	731
QY	654	GAAGCTTGTCCCGCAGTGGAGACGGGACTGCCGATTTGGAGCTAAATTTGGGCCA	713
Db	732	CTATGTGCCAGCAGCTCCAAGGATGGCAGTGTGCGGATCTGGGACACAACACTGAGGCCG	791
QY	714	GTGCCTTATGAACATTGCCGGACACACAAAATGCTGTGACAGCAGTGGGTGGAGC	773
Db	792	CTGTGAGCGCATCCTCACGGGACACCCAGTGGTCACTCTCTCCGGTGGGAGGGGA	851
QY	774	GGGCCTTATTATACATCCTCCAAAGATCGCACAGTGAAGATGTGCGGAGCAGCTGATGG	833
Db	852	CGGGCTTCTCTACTCTGCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG	911
QY	834	AATCTTGTGCGCGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACTTCGCTGAGCAC	893
Db	912	TGTGCTGTGCGGACTCTGCAAGGCCACGGCCACTGGGTGAACACCATGGCCCTCAGCAC	971
QY	894	CGATTACGTCCTGCGCAGTGGTCCATTCATCCGTTGAAGGATCGGTCCAAGAGCC--A	950
Db	972	TGACTATGCCCTGCGCAGTGGGGCCTTTGAACCTGCTGAGGCCCTAGTTAATCCCAAGA	1031
QY	951	CCTCAGTTTGAACACTGAGGAATTGACGAATCTGCTTGAAGCGCTACCAGCGCTGTG	1010
Db	1032	CCTCCAAGGATCCTTGCAGGAGTTGAAGGAGAGGGCTCTGAGCCGATACAACTCGTGGC	1091
QY	1011	CCCTGACGAGGTGGAGTCGCTGGTTTCTGTTGCGATGACAAACACCTCTATCTGTGG--	1068
Db	1092	GGCCAGGTCCAGAGAGGCTGGTGTCTGGCTCCGAGGACTTCACTTATTCCTGTGGTC	1151
QY	1069	-CGGAACAACCAAGAGTGGTTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGA	1127
Db	1152	CCCAGCAGAGGACAAAAGCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCA	1211
QY	1128	TGTGAATATTCCCGGATGTAAAGCTAATTGCGTGTCTTCAATTGACAAGTCAAGTGGC	1187
Db	1212	GGTGCT	1271
QY	1188	TCTGTGGCAGCCAGCGATGGTCACTACATGGCCACTTCCGGGGTCACTGTGCGAGCTGT	1247
Db	1272	GCTGTGGGATGGCAGGACGGGCAAGTACCTGGCTTCCCTACGGCGCCACGTGGCTGCCGT	1331





QY	951	CCTCAGTTTGAGCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTG	1010	PR	28-APR-1999;	99US-0131449.
Db	1039	CCTCCAAGGATCCTTGCAGGAGTTGAAGGAGGGCTCTGAGCCGATACAACCTCGTGG	1098	PR	30-APR-1999;	99US-0132048.
QY	1011	CCCTGACGAGGTGGAGTCGCTGGTTTCCCTGTTCGGATGACAACACCCCTCTATCTGTGG--	1068	PR	30-APR-1999;	99US-0132407.
Db	1099	GGGCCAGGGTCCAGAGAGGCTGGTGTCTGGCTCCGAGACTTCACCTTATTCCTGTGGTTC	1158	PR	04-MAY-1999;	99US-0132484.
QY	1069	-CGGAACAACCAAGTGCCTTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGA	1127	PR	05-MAY-1999;	99US-0132485.
Db	1159	CCCAGCAGAGGACAAAAAGCCTCTCACTCGGATGACAGGCACCAAGCTCTCATCAACCA	1218	PR	06-MAY-1999;	99US-0132486.
QY	1128	TGTGAAATATTCGCCGGATGTAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTGG	1187	PR	06-MAY-1999;	99US-0132487.
Db	1219	GGTGTCTTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCTTTGACAAGTCCATCAA	1278	PR	07-MAY-1999;	99US-0132863.
QY	1188	TCTGTGGCAGCCAGCGATGGTCAGTACATGCCACCTTCGGGGTCAATGTGCAGGCTGT	1247	PR	11-MAY-1999;	99US-0134256.
Db	1279	GCTGTGGATGGCAGGACGGCAAGTACCTGGCTTCCCTACGCGCCACGTGGTGGCGGT	1338	PR	14-MAY-1999;	99US-0134218.
QY	1248	TTACACGGTTGCCGTGGTCCGGGACTCCCGCTTGATTGTTCCGGGCAGAAAGACTCAAC	1307	PR	14-MAY-1999;	99US-0134219.
Db	1339	GTACCAGATTGGTGGTCAGCTGACAGTGGGTCTCTGTGACGGCAGCAGTGCACGCAC	1398	PR	14-MAY-1999;	99US-0134221.
QY	1308	TCTAAAAGTATGGAGTGTGCAGACGAAGAACTGGCACAGAGCTGCCGTGGACATGCGGA	1367	PR	14-MAY-1999;	99US-0134370.
Db	1399	ACTGAAGTGTGGGATGTGAAGGCCCAGAAAGTGGCCATGACCTGCCCGCCACGCGGA	1458	PR	18-MAY-1999;	99US-0134768.
QY	1368	TGAGGTCTTCGGAGTGGACTGGGCGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGA	1427	PR	19-MAY-1999;	99US-0134941.
Db	1459	TGAGGTATATCTGTGTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA	1518	PR	20-MAY-1999;	99US-0135124.
QY	1428	CAAAGTTATAAAGCTATGG	1446	PR	21-MAY-1999;	99US-0135353.
Db	1519	CAAATGCCTCCGGATATGG	1537	PR	24-MAY-1999;	99US-0135629.
RESULT 11						
AAC48725						
ID	AAC48725 standard; DNA; 1726 BP.					
XX	AAC48725;					
AC	18-OCT-2000 (first entry)					
XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 58533.					
DT	Hybridisation assay; genetic mapping; gene expression control;					
DE	protein identification; signal transduction pathway;					
XX	metabolic pathway; promoter; termination sequence; ss.					
KW	Arabidopsis thaliana.					
KW	EP1033405-A2.					
XX	06-SEP-2000.					
PN	25-FEB-2000; 2000EP-0301439.					
PD	25-FEB-1999; 99US-0121825.					
XX	05-MAR-1999; 99US-0123180.					
PF	09-MAR-1999; 99US-0123548.					
XX	23-MAR-1999; 99US-0125788.					
XX	25-MAR-1999; 99US-0126264.					
OS	29-MAR-1999; 99US-0126785.					
XX	01-APR-1999; 99US-0127462.					
XX	06-APR-1999; 99US-0128234.					
PN	08-APR-1999; 99US-0128714.					
XX	16-APR-1999; 99US-0129845.					
XX	19-APR-1999; 99US-0130077.					
XX	21-APR-1999; 99US-0130449.					
XX	23-APR-1999; 99US-0130510.					
XX	23-APR-1999; 99US-0130891.					







PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 21.3%; Score 331.8; DB 21; Length 1643;		
Best Local Similarity 54.3%; Pred. No. 5e-65;		
Matches 722; Conservative 0; Mismatches 592; Indels 15; Gaps 2;		
Qy	129	GCAATTGGGACTGATTGCAACGGCTGCTGATAAAGAGAGCCCTGGAGGACACGTTGGACCTTGGCGTCAGT 188
Db	163	GCAACTTACTCAGCTCGTCAATAGGTTCTTCGACACGAGGAGATGTTACCTTACAGTTT 222
Qy	189	TTTCGTGGCGGAGGATGAGATCAAGAGAGCCCTGGAGGACACGTTGGACCTTGGCGTCAGT 248
Db	223	CTATGTATCAGACGCAAGAGCTTCTTGTACCAGTTGGAAAGCTACTTGGAGAAAAACAAAGT 282
Qy	249	GGACACCGAAAAACGTGATCGATATTGTGTATCAGCCACAGCGGGTTTCAAAGTGGCGCC 308
Db	283	GTCGTGGAGAAAGTTTACGAGATTGTTTATCAACAACAAAGCTGTTTTCGAATTCGTCC 342
Qy	309	AGTGACAAGATGCACGAGTTCCATGCCGGGACACGCCGAGGCTGTGGTTTCGTGAATTT 368
Db	343	TGTTAACCGTTGCTCACAGACAAATTGCTGGTCACGCGGAAGCTGTTCTTGTGTTGCT 402
Qy	369	CAGCCCGGATGGTGTCTCATCTCGCCAGTGGAAAGTGGCGACACGCCACAGTCGAGTTGTGGGA 428
Db	403	TAGTCTGTATGGAAAGCAATTAGCTAGTGTTCAGGTGATACTACTGTACAGGCTTTGGGA 462
Qy	429	TCTTAACACAGAGACACCCGACTTCACCTGCACAGGTCAATAAGCAGTGGGTTCTGTGGGT 488
Db	463	TCTCTACACTGAAACTCCATTGTTTACTTGCAAAAGGCCACAAGAAATTGGGTTCTCACAGT 522
Qy	489	ATCCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTTGCAAAAGCGGCTCTATAATCAT 548
Db	523	TGCGTGGTCTCCTGTATGTTAAGCATCTTGTGAGTGGTAGTAAATCCGGTGAATCTGTTG 582
Qy	549	CTGGGACCCGGAGACGGGTCAGAGAAAGGGGCGACCCCTTGAGTGGGCACAAGAAACACAT 608
Db	583	TTGGAATCCAAAGAAAGGAGAACTAGAAGGACGCCACTTACGGGTACAAGAAATGGAT 642
Qy	609	CAACTGCCTCCCTGGGAACCGGTATCATCCGGATCCGGAGTGCAGGAACCTGCTTCGCG 668
Db	643	TACTGGTATCTCGTGGGAACCAAGTCCACCTTAGTTCTCCATGCCGTGCGATTTGTGACTC 702
Qy	669	CAGTGGAGACGGGACTGCGGATTTGGGAGCTAAAATTTGGGCCACAGTGCCTTATGAACAT 728
Db	703	TAGTAAAGATGGGGATGCAAGGATTTGGGATATTACACTGAAAAAATCTATATTGTGCT 762









GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.  
  
OM nucleic - nucleic search, using sw model  
  
Run on: June 2, 2003, 03:06:49 ; Search time 101 Seconds  
(without alignments)  
4721.608 Million cell updates/sec  
  
Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattcccaaaatgcagga.....aaaaaaaaaaaaaaaa 1555  
  
Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
  
Searched: 441362 seqs, 153338381 residues  
  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
  
Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match %	Length DB ID	Description
1	60.2	3.9	2359 1	US-08-188-582-4 Sequence 4, Appli
2	60.2	3.9	2359 1	US-08-646-715-4 Sequence 4, Appli
3	59	3.8	2674 4	US-09-817-180-1 Sequence 1, Appli
4	57.6	3.7	2550 6	5258287-23 Patent No. 5258287
5	57.4	3.7	1221 3	US-08-965-600-2 Sequence 2, Appli
6	57.4	3.7	1221 4	US-09-489-506-2 Sequence 2, Appli
7	56.2	3.6	10136 1	US-08-353-700-2 Sequence 2, Appli
8	56.2	3.6	10136 5	PCT-US95-16216-2 Sequence 2, Appli
9	56	3.6	1872 4	US-09-801-052-1 Sequence 1, Appli
10	55.8	3.6	2085 2	US-08-283-917-8 Sequence 8, Appli
11	55.8	3.6	2085 2	US-08-961-716-8 Sequence 8, Appli
12	55.4	3.6	1736 3	US-09-182-816-22 Sequence 22, Appl
13	55.4	3.6	1736 3	US-09-182-816-24 Sequence 24, Appl
14	55.4	3.6	1736 3	US-09-471-528-22 Sequence 22, Appl
15	55.4	3.6	1736 3	US-09-471-528-24 Sequence 24, Appl
16	55.4	3.6	1736 4	US-09-634-530-22 Sequence 22, Appl
17	55.4	3.6	1736 4	US-09-634-530-24 Sequence 24, Appl
18	55.4	3.6	7218 1	US-08-232-463-14 Sequence 14, Appl
19	55.2	3.5	812 4	US-09-091-097-7 Sequence 7, Appli
20	54.8	3.5	1212 4	US-09-182-145-34 Sequence 34, Appl
21	54.8	3.5	1212 4	US-09-182-145-35 Sequence 35, Appl
22	54.6	3.5	2323 4	US-09-149-476-24 Sequence 24, Appl
23	54.4	3.5	1098 3	US-09-248-335-35 Sequence 35, Appl
24	54.2	3.5	578 4	US-09-602-877A-95 Sequence 95, Appl
25	54.2	3.5	1066 1	US-08-157-101A-4 Sequence 4, Appli
26	53.8	3.5	1117 4	US-09-247-373B-33 Sequence 33, Appl
27	53.6	3.4	6671 1	US-08-280-443-1 Sequence 1, Appli

28	53.6	3.4	6671 1	US-08-457-459-1 Sequence 1, Appli
29	53.6	3.4	6671 1	US-08-555-678-1 Sequence 1, Appli
30	53.6	3.4	6671 5	PCT-US95-02275-1 Sequence 1, Appli
31	53.4	3.4	2184 4	US-08-955-918C-1 Sequence 1, Appli
32	53.4	3.4	2184 4	US-08-697-766A-1 Sequence 1, Appli
33	53.4	3.4	2447 2	US-09-014-969-14 Sequence 14, Appl
34	53.2	3.4	1411 4	US-08-964-127-5 Sequence 5, Appli
35	53.2	3.4	1411 4	US-09-496-692-5 Sequence 5, Appli
36	53.2	3.4	1582 3	US-08-545-196B-10 Sequence 10, Appl
37	53.2	3.4	1582 3	US-08-545-196B-12 Sequence 12, Appl
38	53.2	3.4	3275 4	US-09-370-838-151 Sequence 151, App
39	53	3.4	144 1	US-08-702-344-26 Sequence 26, Appl
40	53	3.4	1378 4	US-09-149-476-208 Sequence 208, App
41	52.8	3.4	3124 4	US-09-734-030-1 Sequence 1, Appli
42	52.6	3.4	1474 4	US-08-821-994-64 Sequence 64, Appl
43	52.6	3.4	1878 4	US-09-465-558-39 Sequence 39, Appl
44	52.6	3.4	2202 4	US-09-465-558-59 Sequence 59, Appl
45	52.2	3.4	240 1	US-08-628-417-6 Sequence 6, Appli

ALIGNMENTS

RESULT 1  
US-08-188-582-4  
; Sequence 4, Application US/08188582  
; Patent No. 5534410  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlact, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS, OF USE  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/188,582  
; FILING DATE: 28-JAN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman, Richard A  
; REGISTRATION NUMBER: 36,627  
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2359 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 49..2160

US-08-188-582-4

Query Match 3.9%; Score 60.2; DB 1; Length 2359;  
Best Local Similarity 46.2%; Pred. No. 6.1e-07;  
Matches 200; Conservative 0; Mismatches 233; Indels 0; Gaps 0;  
QY 1017 CGAGGTGGAGTCGCTGGTTTCCCTGTCGATGACAAACCCCTCTATCTGTGGCGGAACAA 1076  
Db 1428 CGAGATGAACCTGTGCTCTCATGTTCCGAGGACAGCACCAAGGCTGTGCTCTGCT 1487  
QY 1077 CCAGAACAAAGTCGCTGAGCGCATGACAGGGGACAGGTCAGGTCACGATGTGAATA 1136  
Db 1488 CACCTGGTCCTGCGTAGTCACCTACCGGGGACGTTTACCCGGTGTGGGATGTGCT 1547  
QY 1137 TTCGCCGGATGTAAGCTAATTGGCTCTGCTTCATTGACAAGTCAGTGCCTGTGGCG 1196  
Db 1548 TGGCGCGCATGGCTACTATTTTGTCTTGTTCGTACGACAAAAGTCTGCTGTGGGC 1607  
QY 1197 AGCCAGCGATGGTCAGTACATGCGCACCTTCCGGGGTCATGTGCAGGCTGTTACACGGT 1256  
Db 1608 CACGGATTCCAATCAAGCGTTGGCGTATTTCGTGGGTCACTTGTCCGACGTGGATTGT 1667  
QY 1257 TGCCTGGTCGGGACTCCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAAGT 1316  
Db 1668 ACAATTTTCATCCCAATTCCAATTTATGTGGCCACCGGTTCTAGCGATCGCACGGTAC 1727  
QY 1317 ATGGAGTGTGCAGACGAAGAACTGGCAGGAGTGCCTGGACATCGGATGAGGTGT 1376  
Db 1728 GTGGACAACATGACCGGTGAGTCGGGTACGCGGTATGACGGGCACAAAGGATCGGTG 1787  
QY 1377 CGGAGTGGACTGGGGCCCGGATGGCTCTAGAGTTCCTTGTGGCAAGGACAAAGTTAT 1436  
Db 1788 TTCTCTGGCTTCTCCGCTTCCGCGGTATCTGGCCGTATCTGGCTCGGTTTCAGTAGAT 1847  
QY 1437 AAAGCTATGGCT 1449  
Db 1848 CATCATCTGGGAT 1860

RESULT 2

US-08-646-715-4  
; Sequence 4, Application US/08646715  
; Patent No. 5637686  
; GENERAL INFORMATION:  
; APPLICANT: Tjian, Robert  
; APPLICANT: Comai, Lucio  
; APPLICANT: Dynlacht, Brian D.  
; APPLICANT: Hoey, Timothy  
; APPLICANT: Ruppert, Siegfried  
; APPLICANT: Tanese, Naoko  
; APPLICANT: Wang, Edith  
; APPLICANT: Weinzierl, Robert O.J.  
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/646,715  
; FILING DATE: 09-MAY-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/188,582

FILING DATE: 28-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2359 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 49...2160  
US-08-646-715-4  
Query Match 3.9%; Score 60.2; DB 1; Length 2359;  
Best Local Similarity 46.2%; Pred. No. 6.1e-07;  
Matches 200; Conservative 0; Mismatches 233; Indels 0; Gaps 0;  
QY 1017 CGAGGTGGAGTCGCTGGTTTCCCTGTCGATGACAAACCCCTCTATCTGTGGCGGAACAA 1076  
Db 1428 CGAGATGAACCTGTGCTCTCATGTTCCGAGGACAGCACCAAGGCTGTGCTCTGCT 1487  
QY 1077 CCAGAACAAAGTCGCTGAGCGCATGACAGGGGACAGGTCAGGTCACGATGTGAATA 1136  
Db 1488 CACCTGGTCCTGCGTAGTCACCTACCGGGGACGTTTACCCGGTGTGGGATGTGCT 1547  
QY 1137 TTCGCCGGATGTAAGCTAATTGGCTCTGCTTCATTGACAAGTCAGTGCCTGTGGCG 1196  
Db 1548 TGGCGCGCATGGCTACTATTTTGTCTTGTTCGTACGACAAAAGTCTGCTGTGGGC 1607  
QY 1197 AGCCAGCGATGGTCAGTACATGCGCACCTTCCGGGGTCATGTGCAGGCTGTTACACGGT 1256  
Db 1608 CACGGATTCCAATCAAGCGTTGGCGTATTTCGTGGGTCACTTGTCCGACGTGGATTGT 1667  
QY 1257 TGCCTGGTCGGGACTCCCGCTTGATGTTTCCGGCAGCAAGACTCAACTCTAAAGT 1316  
Db 1668 ACAATTTTCATCCCAATTCCAATTTATGTGGCCACCGGTTCTAGCGATCGCACGGTAC 1727  
QY 1317 ATGGAGTGTGCAGACGAAGAACTGGCAGGAGTGCCTGGACATCGGATGAGGTGT 1376  
Db 1728 GTGGACAACATGACCGGTGAGTCGGGTACGCGGTATGACGGGCACAAAGGATCGGTG 1787  
QY 1377 CGGAGTGGACTGGGGCCCGGATGGCTCTAGAGTTCCTTGTGGCAAGGACAAAGTTAT 1436  
Db 1788 TTCTCTGGCTTCTCCGCTTCCGCGGTATCTGGCCGTATCTGGCTCGGTTTCAGTAGAT 1847  
QY 1437 AAAGCTATGGCT 1449  
Db 1848 CATCATCTGGGAT 1860

RESULT 3

US-09-817-180-1  
; Sequence 1, Application US/09817180  
; Patent No. 6340584  
; GENERAL INFORMATION:  
; APPLICANT: Gan, Weiniu et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001183  
; CURRENT APPLICATION NUMBER: US/09/817,180  
; CURRENT FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1

```
; LENGTH: 2674
; TYPE: DNA
; ORGANISM: Human
US-09-817-180-1

Query Match
Best Local Similarity 3.8%; Score 59; DB 4; Length 2674;
Matches 116; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy 1345 CAGGAGCTGCCTGGACATCGGATGAGGTGTCGGAGTGGACGCGCCGCGCTCT 1404
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2439 CAGCATCCACACTGCGCGGCAGGATGCAGCGCCGTCTCTCTGTGCTGCTGCC 2498

Qy 1405 AGAGTTGCCTCTGTGGCAAGGACAAAGTTATAAAGCTATGGGCTTATTAACAATCATT 1464
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2499 AGGCCTCCTCTTCGCGGCAGAAACAATAAACCACTTGTGCCACTGAATAAAAAA 2558

Qy 1465 AACTTGTACACGGTAAGAAATACTTAGGAATAAAGTAAACGTCCTGAGTAAAAAAA 1524
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2559 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2618

Qy 1525 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2619 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2649
```

```
RESULT 4
5258287-23
; Patent No. 5258287
; APPLICANT: BAXTER, ROBERT C.; WOOD, WILLIAM I.
; TITLE OF INVENTION: DNA ENCODING AND METHODS OF PRODUCTION
; OF INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN BP53
; NUMBER OF SEQUENCES: 58
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/171.623
; FILING DATE: 22-MAR-1988
; SEQ ID NO:23:
; LENGTH: 2550
5258287-23
```

```
Query Match
Best Local Similarity 3.7%; Score 57.6; DB 6; Length 2550;
Matches 84; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 1426 GACAAAGTTATAAAGCTATGGGCTTATTAACAATCATTAACTGTACACGGAAGAAA 1485
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2423 GAGAACTTGTCTTCAATGTATTATAAATAGTAAATAGTTTACCATTAAAAA 2482

Qy 1486 TACTTAGGAATAAAGTAAACGTCCTGAGTAAAAAATAAAAAAAAAAAAAAAAAA 1545
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2483 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2542

Qy 1546 AAAAAA 1553
    |||||
Db 2543 AAAAAA 2550
```

```
RESULT 5
US-08-965-600-2
; Sequence 2, Application US/08965600
; Patent No. 6077688
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW TRANSDUCIN BETA-1 SUBUNIT
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
```

```
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/965,600
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0416 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1221 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KIDNNOT02
; CLONE: 194046
US-08-965-600-2
```

```
Query Match
Best Local Similarity 3.7%; Score 57.4; DB 3; Length 1221;
Matches 157; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

Qy 1130 TGAATATTCGCCGGATGTAAGCTAAATGCGTCTGCTTCATTGACAAGTCAGTGCCTC 1189
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 666 TGACCTTTCCCGGACTCCAGCTCCTGTGCTCAGATGATGGCTACATCAAGA 725

Qy 1190 TGTGGGACCCAGCGATGGTCAGTACATGGCCACCTTCCGGGGTCATGTGCAGGCTGTT 1249
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 726 TCTATGATGTACAACATGCCAATTTGGCTGGCACGGCTGAGCGCCATGCCCTCCTCGGTGC 785

Qy 1250 ACACGGTTCCCTGGTCCCGGACTCCCGCTTGATTGTTCCGGCAGCAAGACTCAACTC 1309
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 786 TGAACGTTGCATTCTGCTGATGACACTCACTTTGTTCCAGTTCGTCGACAAAAGTG 845

Qy 1310 TAAAGTATGGAGTGTGCAGACGGAAGAACTGGCACAGGAGTGCCTGGACATGCGGATG 1369
    ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 846 TAAAGTTGGGATGTTGGAACGAGGACTTGTGTTACACCTTCTTTTCATCACCAGATC 905

Qy 1370 AGGTGTTCCGAGTGGACTGGGCGCCCGGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGACA 1429
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 906 AGGTCTGGGAGTAAATAACAATGGAATGTTCAAAATTTGTCTCTGTTGGAGATGACC 965

Qy 1430 AAGTTATAAGCTATGGGCTTAT 1452
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 966 AGGAAATTCACATCTATGATTGT 988
```

```
RESULT 6
US-09-489-506-2
; Sequence 2, Application US/09489506
; Patent No. 6465619
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: NEW TRANSDUCIN BETA-1 SUBUNIT
; NUMBER OF SEQUENCES: 3
```





COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/16216  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/353,700  
FILING DATE: 09-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Reed, Janet E.  
REGISTRATION NUMBER: 36,252  
TELEPHONE: (215) 563-4100  
TELEFAX: (215) 563-4044  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10136 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US95-16216-2

Query Match 3.6%; Score 56.2; DB 5; Length 10136;  
Best Local Similarity 62.4%; Pred. No. 1.7e-05;  
Matches 88; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 1415 CTGGTGGCAAGGACAAAGTTATATAAGCTATGGGCTTATTAAACAAATCATTAACCTGTGACA 1474  
Db 9990 CTGTTAGCATGGCCATTCCTCTACTGCAATGTAAATAGTATATAAAGCTATGTATATAAAGC 10049  
QY 1475 CGGTAAGAAATACTTAGGAATAAAGTAAACGCTCCTGAGTAAAAAATAAAAAA 1534  
Db 10050 TTTTGGTATATGTTACAAATTAATAAGCAAGCACTATATATAAAAAAATAAAAAA 10109  
QY 1535 AAAAAAATAAAAAAATAAAAAA 1555  
Db 10110 AAAAAAATAAAAAAATAAAAAA 10130

RESULT 9  
US-09-801-052-1  
Sequence 1, Application US/09801052  
Patent No. 6368842  
GENERAL INFORMATION:  
APPLICANT: BEASLEY, Ellen  
TITLE OF INVENTION: ISOLATED HUMAN PHOSPHOLIPASE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHOLIPASE  
TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
FILE REFERENCE: CL001045  
CURRENT APPLICATION NUMBER: US/09/801,052  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1872  
TYPE: DNA  
ORGANISM: Human  
US-09-801-052-1

Query Match 3.6%; Score 56; DB 4; Length 1872;  
Best Local Similarity 64.8%; Pred. No. 8e-06;  
Matches 83; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
QY 1428 CAAAGTTATAAAGCTATGGGCTTATTAAACAAATCATTAACCTGTACACGGTAAGAAATA 1487  
Db 1720 CCAGGCAAAAAGGCCAGGTTATATAAAGTAAATAACTTGTCTGTAAAAAATAAAAAA 1779

QY 1488 CTTAGGATATAAGTAAACGCTCCTGAGTAAAAAATAAAAAAATAAAAAAATAAAAAA 1547  
Db 1780 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1839  
QY 1548 AAAAAAATAAAAAA 1555  
Db 1840 AAAAAAATAAAAAA 1847  
RESULT 10  
US-08-283-917-8  
Sequence 8, Application US/08283917  
Patent No. 5849557  
GENERAL INFORMATION:  
APPLICANT: ADACHI, HIDEKI  
APPLICANT: TSUJIMOTO, MASAFUMI  
APPLICANT: INOUE, KEIZO  
APPLICANT: ARAI, HIROYUKI  
TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME  
TITLE OF INVENTION: AND GENE THEREOF  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER &  
ADDRESSER: NEUSTADT, P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/283,917  
FILING DATE: 03-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 209943/1993  
FILING DATE: 03-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5849557man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2292-030-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2085 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE: Bos taurus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 844..2073  
US-08-283-917-8

Query Match 3.6%; Score 55.8; DB 2; Length 2085;  
Best Local Similarity 46.7%; Pred. No. 9.6e-06;  
Matches 177; Conservative 0; Mismatches 202; Indels 0; Gaps 0;  
QY 1025 AGTCGCTGGTTTCCTGTCGGATGACAAACACCCCTCTATCTGTGGCGGAACACCAACA 1084  
Db 1331 AGCTTCTGGCTTCATGTTCTGTCAGATATGACCATTAAGCTATGGGATTTTCAGGGCTTTG 1390  
QY 1085 AGTCGCTTGAGCGCATGACAGGGGCACAGAACGTCGTCACGATGTGAAATATTTCGCCGG 1144

```

;
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 844..2073
US-08-961-716-8

Query Match      3.6%; Score 55.8; DB 2; Length 2085;
Best Local Similarity 46.7%; Pred. No. 9.6e-06;
Matches 177; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 1025 AGTCGCTGGTTTCCTGTTCCGATGACAAACACCCCTCTATCTGTGGCGGAACAACCAACA 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1331 AGCTTCGGCTTCATGTTCTGCAGATATGACCATTAAGCTATGGGATTTTCAGGGCTTTG 1390

QY 1085 AGTGCCTTGAGCGCATGACAGGGCACCAGAACCTGGTCAACGATGTGAAATATTCGCCGG 1144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1391 AATGCATCAGAACCATGCATGGCCATGACCACATGTTTCTTCAGTAGCCATCATGCCCA 1450

QY 1145 ATGTAAGCTAATTGGCTCTGCTTCATTTGACAAGTCAGTGGCTCTGTGGCGAGCCAGCG 1204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1451 ATGGAGATCATATAGTGTCTGCCCTCAAGGGATAAAACTATATAAAATGTGGGAAGTGCAA 1510

QY 1205 ATGGTCAGTACATGGCCACCTCCGGGGTFCATGTGCAGGCTGTTTACACACGGTTGCTGGT 1264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1511 CTGGCTACTGTGTGAAGACATTCACAGGACACAGAGATGGGTACGTATGGTGGGCCAA 1570

QY 1265 CC CGCGGACTCCCGCTTGATGTTTCCGGCAGCAAAGACTCAACTCTAAAAAGTATGGAGTG 1324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1571 ATCAAGACGGGACTCTGTATAGCCAGCTGTTCCTCAATGACCAGACTGTGCGTGTATGGGTCG 1630

QY 1325 TGCAGACGAAGAACTGGCAGAGGAGCTGCCCTGGACATGCGGATGAGGTGTTCCGGAGTG 1384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1631 TAGCAACAAGGAATGCAAGCTGAGCTTCGAGAACATGAGCATGTGGTAGAATGCATTT 1690

QY 1385 ACTGGGGCGCCGATGGCTC 1403
    ||||| ||| ||| ||| |||
Db 1691 CCTGGGCTCCTGAAAGCTC 1709
```

```

RESULT 11
US-08-961-716-8
; Sequence 8, Application US/08961716
; Patent No. 5880272
; GENERAL INFORMATION:
; APPLICANT: ADACHI, HIDEKI
; APPLICANT: TSUJIMOTO, MASAFUMI
; APPLICANT: INOUE, KEIZO
; APPLICANT: ARAI, HIROYUKI
; TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME
; TITLE OF INVENTION: AND GENE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &
; ADDRESSEE: NEUSTADT,P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,716
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/283,917
; FILING DATE: 03-AUG-1994
; APPLICATION NUMBER: JP 209943/1993
; FILING DATE: 03-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5880272man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2292-030-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
```

```

;
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 844..2073
US-08-961-716-8

Query Match      3.6%; Score 55.8; DB 2; Length 2085;
Best Local Similarity 46.7%; Pred. No. 9.6e-06;
Matches 177; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 1025 AGTCGCTGGTTTCCTGTTCCGATGACAAACACCCCTCTATCTGTGGCGGAACAACA 1084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1331 AGCTTCGGCTTCATGTTCTGCAGATATGACCATTAAGCTATGGGATTTTCAGGGCTTTG 1390

QY 1085 AGTGCCTTGAGCGCATGACAGGGCACCAGAACCTGGTCAACGATGTGAAATATTCGCCGG 1144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1391 AATGCATCAGAACCATGCATGGCCATGACCACATGTTTCTTCAGTAGCCATCATGCCCA 1450

QY 1145 ATGTAAGCTAATTGGCTCTGCTTCATTTGACAAGTCAGTGGCTCTGTGGCGAGCCAGCG 1204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1451 ATGGAGATCATATAGTGTCTGCCCTCAAGGGATAAAACTATATAAAATGTGGGAAGTGCAA 1510

QY 1205 ATGGTCAGTACATGGCCACCTCCGGGGTFCATGTGCAGGCTGTTTACACGCTTGCCTGGT 1264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1511 CTGGCTACTGTGTGAAGACATTCACAGGACACAGAGATGGGTACGTATGGTGGGCCAA 1570

QY 1265 CC CGCGGACTCCCGCTTGATGTTTCCGGCAGCAAAGACTCAACTCTAAAAAGTATGGAGTG 1324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1571 ATCAAGACGGGACTCTGTATAGCCAGCTGTTCCTCAATGACCAGACTGTGCGTGTATGGGTCG 1630

QY 1325 TGCAGACGAAGAACTGGCAGAGGAGCTGCCCTGGACATGCGGATGAGGTGTTCCGGAGTG 1384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1631 TAGCAACAAGGAATGCAAGCTGAGCTTCGAGAACATGAGCATGTGGTAGAATGCATTT 1690

QY 1385 ACTGGGGCGCCGATGGCTC 1403
    ||||| ||| ||| ||| |||
Db 1691 CCTGGGCTCCTGAAAGCTC 1709

RESULT 12
US-09-182-816-22
; Sequence 22, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-182-816-22

Query Match      3.6%; Score 55.4; DB 3; Length 1736;
Best Local Similarity 64.3%; Pred. No. 1.1e-05;
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1427 ACAAACTTATAAGCTATGGCTTATTAAACAATCATTAACGTAAGAAAT 1486
    | || |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 1588 ATAATGTTAAAAATAAATGTAATTACTGTGAAATAAACGATATGGATTTATTTCAAACT 1647  
QY 1487 ACTTAGGAATAAAGTAAACGCTCCTGAGTAAAAAATAAATAAATAAATAAATAAATAA 1546  
Db 1648 TGTCAAATATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1707  
QY 1547 AAAAAAAA 1555  
Db 1708 AAAAAAAA 1716

RESULT 13  
US-09-182-816-24/c  
; Sequence 24, Application US/09182816  
; Patent No. 6143542  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,  
; TITLE OF INVENTION: PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1  
; CURRENT APPLICATION NUMBER: US/09/182,816  
; CURRENT FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-182-816-24

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAAGCTATGGCTTATTAAACAATCATTAACGTGTACACGGTAAGAAAT 1486  
Db 149 ATAATGTTAAAAATAAATGTAATTACTGTGAAATAAACGATATGGATTTATTTCAAACT 90  
QY 1487 ACTTAGGAATAAAGTAAACGCTCCTGAGTAAAAAATAAATAAATAAATAAATAA 1546  
Db 89 TGTCAAATATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 30  
QY 1547 AAAAAAAA 1555  
Db 29 AAAAAAAA 21

RESULT 14  
US-09-471-528-22  
; Sequence 22, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1736

; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (159)..(1553)  
US-09-471-528-22  
Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAAGCTATGGCTTATTAAACAATCATTAACGTGTACACGGTAAGAAAT 1486  
Db 1588 ATAATGTTAAAAATAAATGTAATTACTGTGAAATAAACGATATGGATTTATTTCAAACT 1647  
QY 1487 ACTTAGGAATAAAGTAAACGCTCCTGAGTAAAAAATAAATAAATAAATAAATAA 1546  
Db 1648 TGTCAAATATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1707  
QY 1547 AAAAAAAA 1555  
Db 1708 AAAAAAAA 1716

RESULT 15  
US-09-471-528-24/c  
; Sequence 24, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-471-528-24

Query Match 3.6%; Score 55.4; DB 3; Length 1736;  
Best Local Similarity 64.3%; Pred. No. 1.1e-05;  
Matches 83; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 1427 ACAAGTTATAAAGCTATGGCTTATTAAACAATCATTAACGTGTACACGGTAAGAAAT 1486  
Db 149 ATAATGTTAAAAATAAATGTAATTACTGTGAAATAAACGATATGGATTTATTTCAAACT 90  
QY 1487 ACTTAGGAATAAAGTAAACGCTCCTGAGTAAAAAATAAATAAATAAATAAATAA 1546  
Db 89 TGTCAAATATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 30  
QY 1547 AAAAAAAA 1555  
Db 29 AAAAAAAA 21

Search completed: June 2, 2003, 04:56:39  
Job time : 104 secs







Db 190 CCTGCCACTGGCTTTCTTTGTCACGATGCTGAGATCGTCTCCTCACTGGGAAAGACGTT 249  
QY 234 GGACTTGGCGTCAGTGGACACCCGAAACGTCGATATTTGTGATATCAGCCACAGGCGGT 293  
Db 250 GGAGTCCCAGGAGTGAGACAGAGAGGTCCTAGACATCATCTACCAGCCACAGGCTAT 309  
QY 294 TTTCAAAGTGGCGCCAGTGACAAAGATGCACGAGTTCCTATCGCGGGACACGCCGAGGCTGT 353  
Db 310 CTTCAGAGTCCGGGCTGTGACTCGCTGCACCCAGTCTCTTGAGGGTCCACAGTGAGGCAGT 369  
QY 354 GGTTCGCTGAATTTAGCCCGGATGGTGTCTATCTCGCCAGTGGAAAGTGGCGACACCCAC 413  
Db 370 CATTTCTGTGGCCTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACCCAC 429  
QY 414 AGTGGGATTTGGGATCTTAACACAGACACCCGCCTTACCTGACAGGTTCATAAGCA 473  
Db 430 CGTGGCTTCTGGGATCTCAGCACAGACACACCATTTTCATGCAAGGACACAGACA 489  
QY 474 GTGGTTCTGTGGTATCCTGGGCTCCGGATGGCAACCGTTGGCCAGCGGTTGCAAGC 533  
Db 490 CTGGTCTCTTAGTATATCCTGTCTCCAGATGGCAAGACCTGGCCTCAGGCTGCAAGAA 549  
QY 534 GGGCTCTATAATCATCTGGGACCCGGAGACGGGTGACAGAGGGGGCGACCCCTTGAGTG 593  
Db 550 TGGCCAGATTTCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGCGAGGACCCCTCGCTGG 609  
QY 594 GCACAAGAAACACATCACTGCCTCGCTGGGAAACCGTATCATCGCGATCCGGAGTGCAG 653  
Db 610 CCACAGCAAGTGGATCACAGGCTGAGCTGGGAGCCCTCCATGCGAACCTTGAGTGGCG 669  
QY 654 GAAACTTGTCTCCGCTGAGAGACGGGACTGCCGGATTTGGACGTAATAATTTGGSCCA 713  
Db 670 CTATGTGGCCAGCAGCTCCAAGGATGGCAGTGTGCGGATCTGGGACACAACTGCAGGCCG 729  
QY 714 GTGCCCTTATGAACATTTGCCGACACACAAATGCTGTGACAGCAGTGAGATGGGTTGAGC 773  
Db 730 CTGTGAGCGCATCTCACCGGACACACCCAGTCAGTGTCTCCGGTGGGAGGGGA 789  
QY 774 GGGCCTTATTAACATCTCTCAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG 833  
Db 790 CGGGCTTCTACTCTGCTCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG 849  
QY 834 AATCTTGTCCGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAAATTCGCTGAGCAC 893  
Db 850 TGTGCTGTGCCGACTCTGCAAGGCCACCGCCACTGGGTGAACACCATGGCCCTCAGCAC 909  
QY 894 CGATTACGTCCTGCGCACTGGTCCATTTCCATCCGGTGAAGGATCGGTCCAAGAGCC--A 950  
Db 910 TGACTATGCCCTGCGCACTGGGGCTTTGAACCTGCTGAGGCCCTAGTTAATCCCCAA 969  
QY 951 CCTCAGTTGAGCACTGAGGAATTGCAAGGAATCTGCCTTGAAGCGCTACCAGGCGGTGTG 1010  
Db 970 CCTCCAAGGATCCTTGAGGAGTTGAGGAGAGGGGCTCTGAGCCGATACAAACCTCGTGCG 1029  
QY 1011 CCCTGACGAGGTGGAGTCGCTGTTTCCTGTTCGATGACAAACACCCTCTATCTGTGG-- 1068  
Db 1030 GGGCCAGGTCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACCTTATTCCTGTGGTC 1089  
QY 1069 -CGGAACAACCAAGTGCCTTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGA 1127  
Db 1090 CCCAGCAGAGGACAAAAGCCCTCTCACTCGGATGACAGGACCAACCAAGCTCTCATCAAC 1149  
QY 1128 TGTGAAATATTCGCGGATGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAAGTGG 1187  
Db 1150 GGTGNTCTTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCTTTGACAAGTCCATCAA 1209  
QY 1188 TCTGTGGCGAGCCAGCGATGGTCAATATGCGCACCTTCCGGGGTCAATGTCAGGCTGT 1247  
Db 1210 GCTGTGGGATGGCAGGACGGCAAGTACCTGCTTCCCTACGCGGCCACGTGGCTGCCGT 1269  
QY 1248 TTACACGGTTGCCGTGTCGCGACTCCCGCTTGATTTGTTCCGGCAGCAAGACTCAAC 1307

Db 1270 GTACCAGATTGCGTGGTCACTGACAGTGGCTCTGGTCAAGCGGAGAGTGCAGCAGCAC 1329  
QY 1308 TCTAAAAGTATGGAGTGTGCAGACGAAGAAACTGGCACAGGAGCTGCCTGGACATGCGGA 1367  
Db 1330 ACTGAAGGTGTGGGATGTGAAGGCCCAAGAGCTGGCCATGGACCTGCCCGCCACGCGGA 1389  
QY 1368 TGAGTGTTCGGAGTGGACTGGGCCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGA 1427  
Db 1390 TGAGGTATATGCTGTGACTGGAGTCCAGATGGCCAGAGAGTGGCAAGTGGTGGGAAGGA 1449  
QY 1428 CAAAGTTATAAAGCTATGG 1446  
Db 1450 CAAATGCCCTCCGGATATGG 1468

RESULT 2  
US-10-132-744A-1  
; Sequence 1, Application US/10132744A  
; Publication No. US20030027261A1  
; GENERAL INFORMATION:  
; APPLICANT: Utku, Nalan  
; TITLE OF INVENTION: No. US20030027261A1 genes Tzap7/A, Tzap7/B and Tzap7 involved  
; FILE REFERENCE: Utku-4 CON  
; CURRENT APPLICATION NUMBER: US/10/132,744A  
; CURRENT FILING DATE: 2002-04-24  
; PRIOR APPLICATION NUMBER: PCT/EP00/10670  
; PRIOR FILING DATE: 2000-10-30  
; PRIOR APPLICATION NUMBER: 60/185,016  
; PRIOR FILING DATE: 2000-02-25  
; PRIOR APPLICATION NUMBER: 60/162,675  
; PRIOR FILING DATE: 1999-11-01  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1928  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (92)..(1543)  
US-10-132-744A-1

Query Match 30.0%; Score 466.2; DB 9; Length 1928;  
Best Local Similarity 59.3%; Pred. No. 5.le-115;  
Matches 830; Conservative 0; Mismatches 563; Indels 6; Gaps 2;

QY 54 GATACAGGCGCGCTCGTTTACACGGGCGAGGAAGCCGCCCAATCGACCTGCCGCG 113  
Db 139 GCTAGTGCAGTTCAGGATGAGGGCGGCGACGCTGCTGGGTTCCCCGTTGACGCTGCCGCT 198  
QY 114 AGGAATCACTACCCAGCAATTTGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGC 173  
Db 199 GGACATCACCCCGGACAGGCTGCAGCTCGTTTGCAACGCGCTACTTGGCCCGAGGAGATCC 258  
QY 174 CACTCCATATTTGTTTTTCGTGGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACAGCTT 233  
Db 259 CCTGTCACTGGCTTTCTTTGTGCCAGATGCTGAGATCGTTTCCCTCACTCGGGAAGAGCTT 318  
QY 234 GGACTTGGCGTCAGTGGACACCCGAAACGTCGATTCGATATTTGTATCAGCCACAGGCGGT 293  
Db 319 GGAGTCCAGGCGAGTGGAGACAGAGAAGGTCCTAGACATCATTTTACCCACACAAGCTGT 378  
QY 294 TTTCAAAGTGGCCCCAGTGACAAAGATGCACAGTTCCATGCCGGGACACGCCGAGGCTGT 353  
Db 379 GTTCAAAGTTCGTGCTGAACACGATGTACCAAGTATTTGGCAAGTGGTTCTGGGACACTAC 438  
QY 354 GGTTCGCTGAATTTACGCCCGGATGGTGTCTCATCTGCCAGTGGGAAGTGGCGACACCCAC 413  
Db 439 TATTCAGTAGCCTTCAGCCCCAACTGGAAAAGTATTTGGCAAGTGGTTCTGGGACACTAC 498  
QY 414 AGTGCATTTGGGATCTTAACACAGAGACACCGCACTTTCACCTGCACAGGTCATAAGCA 473

499	AGTCCGCTTTTGGGATCTCAGCAGACAGAAACTCCACATTTTACATCTAAAGGGCATACACA	558
474	GTGGGTTCTGTGCGTATCCTGGGCTCCGGATGGGCAAAACGGTTGGCCACGCGTTGCAAAAGC	533
559	CTGGGTTCTCAGTATTGCTTGGTCTCCAGATGGCAAAAACCTTGCCCTCAGGATGTAAAAA	618
534	GGGCTCTAATAATCATCTGGGACCCGGAGACGGGTACGACAGAAGGGCGACCCCTTGAGTGG	593
619	TAGTCAGATCTTCATTTGGGACCCCAAGCACAGGGAAGCAGATTTGGCAACCATTAACAGG	678
594	GCACAAGAAACACATCAACTGCTTCGCCCTGGGAAACCGTATCATCGCGATCGGAGTGCAG	653
679	GCACCTCAAAGTGGATTACATTTGGTGTGTGGGAACCTCTCCACCTGAACCCAGAGAGCCG	738
654	GAACACTTGCTTCCGCCAGTGGAGACGGGACTCCGGATTTGGGACGTAATAATTGGGCCA	713
739	ATACCTAGCCAGTGCCTCCAGCGGCCGCGTCGACCCGATCTGGGACACAACCTGCAGGCCG	798
714	GTGCCTTATGAACATTTGCCGACACACAATGCTGTGACAGCAGTGAGATGGGTGGAGC	773
799	CTGTGAGCGCATCTCACCGGGCACACCCAGTCGGTCACTGTCTCCGGTGGGAGGGGA	858
774	GGGCCTTATTATACATCTCTCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG	833
859	CGGGCTTCTACTCTGCTCCACAGGACCCGACCATCAAAGTCTGGAGAGCTCATGACGG	918
834	AATCTTGTGGCGGACGTTCTCTGGCCCAAGCTCACTGGGTAACAACATTCGCTGAGCAC	893
919	TGTGCTGTGCGGACTCTGCAAGGCCACGGCCACTGGGTGAACACCATGGCCCTCAGCAC	978
894	CGATTACGTCCTGCGCACTGCTGCCATTCATTCGGGTGAAGGATCGCTCCAAGAGCC--A	950
979	TGACTATGCCCTCGCACTGGGGCCCTTTGAACTGCTGAGGCTCAGTTAATCCCAAGA	1038
951	CCTCAGTTTGAGCACTGAGGAATTCAGGAATCTGCCTTGAAGCGCTACCAGGCCGTGTG	1010
1039	CCTCCAAGGATCTTGACGAGTGAAGGAGAGGGCTCTGAGCCGATACAACTCGTGCG	1098
1011	CCCTGACGAGTGGAGTCGCTGGTTTCTGTTTCGGATGACAAACCCCTCTATCTGTGG--	1068
1099	GGGCCAGGGTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTCACCTTATTCCTGTGGTC	1158
1069	-CGGAACAACAGAAAGTGGTTGAGCGCATGACAGGGCACAGAACTGCTCAACGA	1127
1159	CCCAGCAGAGACAAAAGCCCTCTCACTCGGATGACAGGACACAAAGCTCTCATCAACCA	1218
1128	TGTGAAATATTCCCGGATGFAAAGCTAATTGGCTCTGCTTCATTTGACAAGTCAGTGC	1187
1219	GGTGCTCTTCTCTGACTCCCGCATCGTGGCTAGTGCCTCTCTTGACAAGTCCATCAA	1278
1188	TCTGTGGCGAGCCAGCATGGTFCAGTACATGGCCACCTTCGGGGTTCATGTGAGGGCTGT	1247
1279	GCTGTGGGATGGCAGGACGGGCAAGTACCTGGCTTCCCTACGCGGCCACGCTGCTGCCGT	1338
1248	TTACACGGTTGCCTGGTCCGGGACTCCCGCTGATTTTCCGGCAGCAAAAGACTCAAC	1307
1339	GTACCAGATTGGTGGTCAAGTACAGTTCGGCTCTGCTGTCAGCGCGCAGCAGTGACAGCAC	1398
1308	TCTAAAAGTATGGAGTGTGCAGACGAAGAACTGGCACAGGAGCTGCCTGGACATGCGGA	1367
1399	ACTGAAGGTGTGGGATGTGAAGGCCCCAGAAAGTGGCCATGGACCTGCCCGGCCACGCGGA	1458
1368	TGAGGTGTTCCGAGTGGACTGGGCGCCCGATGGCTCTAGAGTTGCCCTCTGTTGGCAAGGA	1427
1459	TGAGGTATATGCTGTTGACTGGAGTCCAGATGGCCATGGACATGGCAAGTGGTGGGGAAGGA	1518
1428	CAAAGTTATAAGCTATGG	1446
1519	CAAATGCCCTCCGGATATGG	1537

```

; Sequence 3, Application US/10132744A
; Publication No. US20030027261A1
; GENERAL INFORMATION:
; APPLICANT: Utku, Nalan
; TITLE OF INVENTION: No. US20030027261A1el genes Tzap7/A, Tzap7/B and Tzap7 i
; TITLE OF INVENTION: activation and uses thereof
; FILE REFERENCE: Utku-4 CON
; CURRENT APPLICATION NUMBER: US/10/132,744A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: PCT/EP00/10670
; PRIOR FILING DATE: 2000-10-30
; PRIOR APPLICATION NUMBER: 60/185,016
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 60/162,675
; PRIOR FILING DATE: 1999-11-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(785)
;
US-10-132-744A-3

```

Query Match	16.0%;	Score 249.4;	DB 9;	Length 1170;
Best Local Similarity	59.3%;	Pred. No. 7.6e-57;		
Matches 462; Conservative	0;	Mismatches 311;	Indels 6;	Gaps 2;

QY	674	GAGACGGGAGCTCCGGATTGGGACGTAATAATTGGCCAGTGCCTTTATGAACATTGCCG	733
Db	1	CGGGCCGCTGACCGGATCTGGGACACAACTGCAGGCCGCTGTGAGCGCATCCTCACCG	60
QY	734	GACACACAATGCTGTGACAGCAGTGAGATGGGGTGGAGCGGGCCTTTATTATACATCCT	793
Db	61	GGCACACCCAGTCGGTACCTGTCTCCGGTGGGGAGGGGACGGGCTTCTTACTCTGCCT	120
QY	794	CCAAAGATCCGACAGTGAAGATGTGGCAGCAGCTGATGGAACTCTTGTGCGCGGACGTTCT	853
Db	121	CCCAGGACCCGACCAATCAAAGTCTGGAGAGTCTATGACGGTGTGCTGTGCGGACTCTGC	180
QY	854	CTGGCCAAAGTCACTGGGTAAACAACATTCGGCTGAGCACCGGATTACGTCCTGCGCACGTG	913
Db	181	AAGGCCACGGCCACTGGGTGAACACCATTGGCCCTCAGCACTGACTATGCCCTGCGCACTG	240
QY	914	GTCCATTCCATCCGGTGAAGATCGCTCCAAGAGCC--ACCTCAGTTTGAGCACTGAGG	970
Db	241	GGGCCTTTGAACCTGCTGAGGCCCTCAGTTAATCCCCAAGACCTCCAAGGATCCTTGCAGG	300
QY	971	AATTGCAGGAATCTGCCCTTGAGCGCTACCGAGCCCGTGTGCCCTGACGAGGTGGAGTCGC	1030
Db	301	AGTTGAAGGAGAGGGCTCTGAGCCGATACAACCTCGTGGCGGGCCAGGGTCCAGAGAGGC	360
QY	1031	TGGTTTCCTGTTCGGATGACAACACCCCTCTATCTGTGG--CGGAACAACCAGAACAAAGT	1087
Db	361	TGGTGTCTGGCTCCGACGACTTCACTTATTCCTGTGGTCCCGCAGCAGAGACACAAAAGC	420
QY	1088	GCGTTGAGCGGATGACAGGACCCAGAACCTGGGTCAACGATGTGAAATATTCGCCGGATG	1147
Db	421	CTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCAGGTGCTCTTCTCTCCTGACT	480
QY	1148	TAAAGCTAATTGCGTCTGCTTCAATTTGACAAGTCAGTGCCTGTGTGGCGAGCCAGCGATG	1207
Db	481	CCCGCATCGTGGCTAGTGCCTCCTTTTGACAAGTCCATCAAGCTGTGGGATGGCAGGACGG	540
QY	1208	GTCAGTACATGGCCACCTTCGGGGGTCAATGTCAGGCTGTTTACACGGTTGCCCTGGTCCG	1267
Db	541	GCAAGTACCTGGCTTCCCTACGCGGCCACCTGGCTGCCGGTACCAAGATTGCGTGGTCTAG	600
QY	1268	CGGACTCCCGCTGATTGTTTCCGGCAGCAAAAGCACTCAACTCTAAAAGTATGGAGTGTGC	1327
Db	601	CTGACAGTCCGCTCCTGGTCAGCGCGGACGAGTGCAGCAGCACTGAAGGTGTGGGATGTGA	660







Db 63 GAGCTCATGACGGTGTGCTGTGCCGGACTCTGCAAGGCCACGCGCACTGGGTGAACACCA 122  
Qy 881 TTGGCGCTGAGCACCGGATTACGTCTCGCGACTGGTCCATTCCATCCGGTGAAGGATCGCT 940  
Db 123 TGGCCCTCAGCACTGACTATGCCCTGCGCACTGGGGCCCTTTGAACCTGCTGAGGCGCTCAG 182  
Qy 941 CCAAGAGCC---ACCTCAGTTTGAGCACTGAGGAATTGCAGGAATCTGCCCTGAAGCGCT 997  
Db 183 TTAATCCCCAAGACCTCCAAGGATCCTTGCAGGAGTTGAAGGAGAGGGCTCTGAGCCGAT 242  
Qy 998 ACCAGGCCGTGTGCCCTGACGAGGTGGAGTCGCTGGTTTCCTGTTTCGGATGACAAACACCC 1057  
Db 243 ACRACCTCGTGGGGGCCAGGGTCCAGAGAGGCTGGTGTCTGGCTCCGACGACTTCACCT 302  
Qy 1058 TCATCTGTGG---CGGAACAACCAAGAGTGCCTTGAGCGCATGACAGGGCACCAAGA 1114  
Db 303 TATTCTGTGGTCCCCCAGCAGGACAAAAGCCCTCTCACTCGGATGACAGGACACCAAG 362  
Qy 1115 ACGTGGTCAACGATGTGAAATATTCCGCCGGATGTAAAGCTAATTGCGTGTGCTTCATTGG 1174  
Db 363 CTCTCATCAACCAAGGTGCTCTCTCTCTGACTCCCGCATCGTGGCTAGTGCCTCCTTTG 422  
Qy 1175 ACAAGTCAGTGGTCTGTGG 1194  
Db 423 ACAAGTCCATCAAGCTGTGG 442

RESULT 8  
US-09-764-853-29  
; Sequence 29, Application US/09764853  
; Patent No. US20020090672A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PJZ06  
; CURRENT APPLICATION NUMBER: US/09/764,853  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 939  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 3025  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (3011)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (3014)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (3015)  
; OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-853-29

Query Match 5.1%; Score 79.4; DB 10; Length 3025;  
Best Local Similarity 49.2%; Pred. No. 6.5e-11;  
Matches 209; Conservative 0; Mismatches 216; Indels 0; Gaps 0;  
Qy 1025 AGTCGCTGGTTTCCTGTCGGATGACAAACCCCTCTATCTGTGGGGGAACACCAAGAACA 1084  
Db 222 AGCAACTTGCTACTGCTTCTTGGGATACCTTTCTCATGCTATGGAATTTCAAGCCACATG 281  
Qy 1085 AGTCGCTTGAGCGCATGACAGGGGACCAGACGTTGGTCAACGATGTGAAATATTCCCGG 1144  
Db 282 CTAGAGCTTACAGATATGTGGTTCACAAGGATGTGTAACCGGTGCAGTTTCTCCAC 341  
Qy 1145 ATGTAAGCTAATTGGTCTGCTTCATTTGACAAAGTCAGTGCCTCTGTGGCGAGCCAGCG 1204  
Db 342 ATGGAACCTTATTGGGCTGCTGCCCTCAGCAGACAGAACCCGTGAGACTCTGGATTCTCTGATA 401

Qy 1205 ATGGTCAGTACATGGCCACCTTCCGGGGTCATGTGCAGGCTGTTTACACGGTTGCCTGGT 1264  
Db 402 AGAGAGGAAAAATTCTCAGAATTTAAAGCTCATACAGCTCCAGTTTGAAGTGTAGACTTTT 461  
Qy 1265 CCGCGACTCCCGCTTGATTGTTTCGGGAGCAAGAGACTCAACTCTAAAAGTATGGAGTG 1324  
Db 462 CAGCTGATGGCCAGTTTCTAGCTACAGCTTCTGAAGACAAATCCATAAAAGTATGGAGCA 521  
Qy 1325 TGCAGACGAAGAACTGGCACAGGAGCTGCCCTGGACATGCGGATGAGGTGTTTCGAGTGG 1384  
Db 522 TGTATGCCAGCGCTTCTGTATTCTTGTATCGACATACACACTGGGTACGCTGTGCCA 581  
Qy 1385 ACTGGGCGCCCGATGGCTCTAGAGTTGCCCTCTGTGGCAAGGACAAAAGTTATAAAGCTAT 1444  
Db 582 AATTTTCAACCCGATGGAAGACTAATTGTGTATGTAGTGAGGATAAAAACCTATTAATAATT 641  
Qy 1445 GGGCT 1449  
Db 642 GGGAT 646

RESULT 9  
US-09-794-257-4  
; Sequence 4, Application US/09794257  
; Patent No. US20020009804A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyers, Rachel  
; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US20020009804A1e1  
; TITLE OF INVENTION: Human G-Proteins  
; FILE REFERENCE: 35800/209285  
; CURRENT APPLICATION NUMBER: US/09/794,257  
; CURRENT FILING DATE: 2001-02-27  
; PRIOR APPLICATION NUMBER: 60/185,606  
; PRIOR FILING DATE: 2000-02-29  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 1023  
; TYPE: DNA  
; ORGANISM: homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (245)...(886)  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1023)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-794-257-4

Query Match 3.8%; Score 59.6; DB 10; Length 1023;  
Best Local Similarity 72.0%; Pred. No. 7.5e-06;  
Matches 77; Conservative 0; Mismatches 30; Indels 0; Gaps 0;  
Qy 1449 TTATTAACAAATCACTTAACCTGTACACGGTAAGAAAAAACTTAGGAATAAAACGT 1508  
Db 896 TTATTTCCAAAACATGCTCTCTACTTGAACCTGAAAAGTAAGAGAAAATAAGATCTT 955  
Qy 1509 CCTGAGTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555  
Db 956 TGTGTNAAA 1002

RESULT 10  
US-09-925-300-464  
; Sequence 464, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988

```
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 464
; LENGTH: 2431
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-464

Query Match      3.8%; Score 59.4; DB 10; Length 2431;
Best Local Similarity 77.4%; Pred. No. 1.4e-05;
Matches 72; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 1463 TTAACCTGTACACGGTAAGAAATACTTAGGAATAAAGTAAACGTCCTGACGTAAAAA 1522
Db 2337 TTGGTTGTATTCTGTAGCTCAGTAGCTGCTAATAAAGTTAAAGATCCTGAAAAA 2396

QY 1523 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555
Db 2397 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2429

RESULT 11
US-10-003-295-1
; Sequence 1, Application US/10003295
; Patent No. US20020168741A1
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: C1001183DIV
; CURRENT APPLICATION NUMBER: US/10/003,295
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2674
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-003-295-1

Query Match      3.8%; Score 59; DB 9; Length 2674;
Best Local Similarity 55.0%; Pred. No. 1.9e-05;
Matches 116; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 1345 CAGGAGTCCTGGACATGCGGATGAGGTGTTCCGAGTGGACTGGCGCCCGATGGCTCT 1404
Db 2439 CAGCATCCACACTGCCGCGAGGATGACGCGCGCTCTCTGTGTCCCTGTGCTGCC 2498

QY 1405 AGAGTTGCTCTGTGGCAAGGACAAAGTTATAAGCTATGGCTTATTAAACAATCATT 1464
Db 2499 AGGGCTTCTCTTCGCGCAGAAACAATAAACCACTTGTGTGCCACTGAAAAA 2558

QY 1465 AACTTGTACACGGTAAGAAATACTTAGGAATAAAGTAAACGTCCTGAGTAAAAA 1524
Db 2559 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2618

QY 1525 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555
Db 2619 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2649

RESULT 12
US-09-918-995-30837
; Sequence 30837, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
```

```
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30837
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(453)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-30837

Query Match      3.8%; Score 58.4; DB 9; Length 453;
Best Local Similarity 48.8%; Pred. No. 1e-05;
Matches 158; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 1019 AGGTGGAGTCGCTGGTTTCCTGTTCCGATGACAAACCCCTCTATCTGTGGCGGAACACC 1078
Db 60 ACGGCGAGTTGTAGCCTCTGTTCCAGGACACAAACATCAAGCTCTGGGACATCAGGA 119

QY 1079 AGAACAAAGTCGTTGAGCCCATGACAGGCGCACAGACCTGGTCAACGATGTGAAATATT 1138
Db 120 GGAAAGGCTGTGCTTCCGATACAGGGGGGCACAGCCAGGCCCTGCGGTCTCTCCGGTTCA 179

QY 1139 GCGCGGATGTAAAGCTAATTGCGTCTGCTTCATTTGACAAGTCAGTGCCTCTGTGGCGAG 1198
Db 180 GCCCGGATGGGAAGTGGTGGCGTCGGCGCAGATGACACACCGTGAAGCTCTGGGATC 239

QY 1199 CCAGCGATGTCAGTACATGCGCCACCTTCCGGGGTTCATGTCAGGCTGTTACACGGTTG 1258
Db 240 TCACTGCCGCAAGATGATGCTCTGAGTTCCTGTGTCACACGGGGCCTGTCAACGTTGGTCG 299

QY 1259 CCTGGTCCGCGGACTCCCGCTGATTTGTTCCGGCAGCAAAAGACTCAACTCTAAAAGTAT 1318
Db 300 AGTTTCAACCCCAACGAGTACTCTCTGCTGCTCCGGCAGCTCTGACAGGACATCCGCTTCT 359

QY 1319 GGAGTGTGCAGACGAAAGAACTGG 1342
Db 360 GGGACCTGGAGAAAGTTCCAGGTGG 383

RESULT 13
US-09-997-279-19
; Sequence 19, Application US/09997279
; Publication No. US20030059781A1
; GENERAL INFORMATION:
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN AND ENDOMETRIAL CANCER
; FILE REFERENCE: 210121.501C1
; CURRENT APPLICATION NUMBER: US/09/997,279
; CURRENT FILING DATE: 2001-11-28
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 1602
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1602)
; OTHER INFORMATION: n = A,T,C or G
US-09-997-279-19

Query Match      3.7%; Score 57.6; DB 9; Length 1602;
Best Local Similarity 65.1%; Pred. No. 3.3e-05;
Matches 84; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 2, 2003, 02:57:24 ; Search time 2084 Seconds  
(without alignments)  
12084.444 Million cell updates/sec

Title: US-09-830-980-2  
Perfect score: 1555  
Sequence: 1 aattcccaaaaaatgcagga.....aaaaaaaaaaaaaaaaaaaa 1555

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :		EST:*	
1:	em_estba:*	1:	em_estba:*
2:	em_esthum:*	2:	em_esthum:*
3:	em_estin:*	3:	em_estin:*
4:	em_estnu:*	4:	em_estnu:*
5:	em_estov:*	5:	em_estov:*
6:	em_estpl:*	6:	em_estpl:*
7:	em_estro:*	7:	em_estro:*
8:	em_htc:*	8:	em_htc:*
9:	gb_est1:*	9:	gb_est1:*
10:	gb_est2:*	10:	gb_est2:*
11:	gb_htc:*	11:	gb_htc:*
12:	gb_est3:*	12:	gb_est3:*
13:	gb_est4:*	13:	gb_est4:*
14:	gb_est5:*	14:	gb_est5:*
15:	em_estfun:*	15:	em_estfun:*
16:	em_estom:*	16:	em_estom:*
17:	gb_gss:*	17:	gb_gss:*
18:	em_gss_hum:*	18:	em_gss_hum:*
19:	em_gss_inv:*	19:	em_gss_inv:*
20:	em_gss_pln:*	20:	em_gss_pln:*
21:	em_gss_vrt:*	21:	em_gss_vrt:*
22:	em_gss_fun:*	22:	em_gss_fun:*
23:	em_gss_mam:*	23:	em_gss_mam:*
24:	em_gss_mus:*	24:	em_gss_mus:*
25:	em_gss_other:*	25:	em_gss_other:*
26:	em_gss_pro:*	26:	em_gss_pro:*
27:	em_gss_rod:*	27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	735.6	47.3	790	9 AI518473	AI518473 LD38001.5
2	597.8	38.4	637	13 BI637470	BI637470 SD19506.5
3	591.8	38.1	725	12 BF505998	BF505998 AT08344.5
4	578.2	37.2	622	13 BI632742	BI632742 SD26605.5
5	549.2	35.3	606	13 BI232464	BI232464 RE28535.5
6	548.8	35.3	626	13 BI639652	BI639652 SD22291.5

7	488.2	31.4	527	9	AI520108	AI520108 LD40082.5
8	485.2	31.2	552	13	BI635777	BI635777 SD17257.5
9	476.6	30.6	1784	11	BC035421	BC035421 Mus muscu
10	455.8	29.3	510	13	BI635228	BI635228 SD16515.5
11	424	27.3	448	9	AA735870	AA735870 GM10004.5
12	346	22.3	688	13	BM579586	BM579586 170006872
13	331.6	21.3	975	9	AL520339	AL520339 AL520339
14	314	20.2	925	9	AL526575	AL526575 AL526575
15	301.4	19.4	1114	9	AL530778	AL530778 AL530778
16	296.4	19.1	861	9	AL558947	AL558947 AL558947
17	290.4	18.7	851	9	AL558453	AL558453 AL558453
18	289.4	18.6	861	9	AL519702	AL519702 AL519702
19	284.2	18.3	997	14	BQ070427	BQ070427 AGENCOURT
20	282	18.1	927	12	BG470097	BG470097 602533586
21	278.8	17.9	715	13	BM579239	BM579239 170006872
22	278.8	17.9	939	14	BQ956834	BQ956834 AGENCOURT
23	276.6	17.8	826	9	AL526922	AL526922 AL526922
24	275.2	17.7	822	9	AL550330	AL550330 AL550330
25	268.8	17.3	1103	13	BM478420	BM478420 AGENCOURT
26	265.2	17.1	1006	13	BM474111	BM474111 AGENCOURT
27	263.6	17.0	816	9	AL527738	AL527738 AL527738
28	263	16.9	573	13	BM654928	BM654928 170006873
29	262.4	16.9	324	13	BI232484	BI232484 RE28559.5
30	256.6	16.5	921	13	BI833055	BI833055 603090947
31	251.4	16.2	912	9	AL522417	AL522417 AL522417
32	250.2	16.1	903	12	BG481047	BG481047 602529007
33	245.4	15.8	673	13	BM654934	BM654934 170006873
34	242.6	15.6	678	14	BQ257262	BQ257262 NISC_ko12
35	242.6	15.6	776	12	BG323761	BG323761 602421865
36	242	15.6	846	13	BI116063	BI116063 602866459
37	241.2	15.5	1080	13	BM467429	BM467429 AGENCOURT
38	239.4	15.4	756	12	BG386139	BG386139 602455319
39	239	15.4	897	14	BQ727393	BQ727393 AGENCOURT
40	238.4	15.3	697	13	BI855975	BI855975 603383349
41	237.4	15.3	843	12	BE792381	BE792381 601585118
42	236.2	15.2	814	13	BI772615	BI772615 603057607
43	233.6	15.0	759	12	BF536680	BF536680 602048656
44	232.6	15.0	274	9	AA696993	AA696993 GM08866.5
45	229.8	14.8	668	14	BM849233	BM849233 K-EST0129

ALIGNMENTS

RESULT 1	AI518473	790 bp	mRNA	linear	EST 23-APR-2001
LOCUS	LD38001.5	prime LD Drosophila melanogaster	embryo pOT2 Drosophila		
DEFINITION	LD38001.5	melanogaster cDNA clone LD38001	5 similar to Nle: FBan0002863		
		'signal transduction' located on:	2L 21C6-21C6;: 04/10/2001, mRNA		
		sequence.			
ACCESSION	AI518473				
VERSION	AI518473.2	GI:13769352			
KEYWORDS	EST.				
SOURCE	fruit fly.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
	Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 790)				
AUTHORS	Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.				
	BDGP/HMI Drosophila EST Project				
TITLE	Unpublished (2001)				
JOURNAL	On Mar 16, 1999 this sequence version replaced gi:4424327.				
COMMENT	Other_ESTs: LD38001.3prime				
	Contact: Stapleton, M.				
	BDGP				
	Lawrence Berkeley National Lab				
	One Cyclotron Rd, Berkeley, CA 94720, USA				
	Fax: 510 486 6798				
	Email: <a href="http://www.fruitfly.org/EST">http://www.fruitfly.org/EST</a> , <a href="mailto:est@fruitfly.berkeley.edu">est@fruitfly.berkeley.edu</a>				
	hit genomic AE003589: arm:2L [301580,604052]				

estimated-cyto:21B7-21C7: 04/10/2001  
Plate: LD.380 row: A column: 1  
High quality sequence stop: 626  
POLYA-No.

FEATURES  
source

Location/Qualifiers

1..790  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="LD38001"  
/clone\_lib="LD Drosophila melanogaster embryo pot2"  
/sex="male and female"  
/dev\_stage="0 to 24 hours mixed stage embryonic"  
/lab\_host="XLI Blue"  
/note="Organ: embryo; Vector: pot2; Site\_1: EcoRI; Site\_2:  
xhoI; Sized fractionated cDNAs were directly ligated into  
pot2."

BASE COUNT 199 a 198 c 247 g 146 t  
ORIGIN  
Query Match 47.3%; Score 735.6; DB 9; Length 790;  
Best Local Similarity 98.8%; Pred. No. 8.1e-110;  
Matches 763; Conservative 0; Mismatches 4; Indels 5; Gaps 2;

QY 7 CAAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCACATACGATACAGGCGGC 66  
DB 19 CAGAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCACATACGATACAGGCGGC 78  
QY 67 CTCGTTT---ACACGGGCGAGGAAGCCGCCGCCCAATCGACCTGCCGCGAGGAATCACT 123  
DB 79 CTCGTTTCGGACACGGGCGAGGAAGCGGCCGCCCAATCGACCTGCCGCGAGGAATCACT 138  
QY 124 ACCCAGCAATTGGGACTGATTTGCAACGGCTGCTGAAAAACGAGGAAGCCACTCCATAT 183  
DB 139 ACCCAGCAATTGGGCTGATTTGCAACGGCTGCTGAAAAACGAGGAAGCCACTCCATAT 198  
QY 184 TTGTTTTTCGTTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACCTTGACTTGGCG 243  
DB 199 TTGTTTTTCGTTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACCTTGACTTGGCG 258  
QY 244 TCAGTGGACACCGAAAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTG 303  
DB 259 TCAGTGGACACCGAAAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTG 318  
QY 304 CGCCCAGTGACAAGATGACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 363  
DB 319 CGCCCAGTGACAAGATGACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 378  
QY 364 AATTTCAGCCCGGATGGTGCTCATCTCCAGTGGAGTGGCGACAC--CACAGTCCGAT 421  
DB 379 AATTTCAGCCCGGATGGTGCTCATCTCCAGTGGAGTGGCGACACACACAGTCCGAT 438  
QY 422 TGTGGGATCTTAACACAGAGACACCCGCACTTCACCTGCACAGGTCTAAGCAGTGGGTTT 481  
DB 439 TGTGGGATCTTAACACAGAGACACCCGCACTTCACCTGCACAGGTCTAAGCAGTGGGTTT 498  
QY 482 TGTGCGTATCCTGGGCTCCGGATGGCAACCGTTTGGCCAGCGTTTCAAAAGCGGCTCTA 541  
DB 499 TGTGCGTATCCTGGGCTCCGGATGGCAACCGTTTGGCCAGCGTTTCAAAAGCGGCTCTA 558  
QY 542 TAATCATCTGGGACCCGGAGACGGGTGAGAGAGGGGCGGACCCCTTGAGTGGGCACAAGA 601  
DB 559 TAATCATCTGGGACCCGGAGACGGGTGAGAGAGGGGCGGACCCCTTGAGTGGGCACAAGA 618  
QY 602 AACACATCAACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAACCTG 661  
DB 619 AACACATCAACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAACCTG 678  
QY 662 CTTCCGCCAGTGGAGACGGGACTGCGCGATTGGGACGTAAATTTGGCCAGTGCCTTA 721  
DB 679 CTTCCGCCAGTGGAGACGGGACTGCGCGATTGGGACGTAAATTTGGCCAGTGCCTTA 738  
QY 722 TGAACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGGTGGAGC 773

Db 739 TGAACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGTTGGAGC 790

RESULT 2

BI637470

LOCUS

DEFINITION

BI637470 637 bp mRNA linear EST 10-SEP-2001  
SD19506.5prime SD Drosophila melanogaster Schneider L2 cell culture  
pot2 Drosophila melanogaster cDNA clone SD19506 5 similar to Nle:  
FBan0002863 GO:[signal transduction (GO:0004871)] located on: 2L  
21C6-21C6;: 05/19/2001, mRNA sequence.

ACCESSION

VERSION BI637470.1 GI:15539680

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster

fruit fly.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

1 (bases 1 to 637)  
Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,  
Lewis,S. and Rubin,G.M.

TITLE

JOURNAL

COMMENT

BDGP/HIMI Drosophila EST Project  
Unpublished (2001)  
Contact: Stapleton, M.  
BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)

hit genomic AE003589: arm:2L [301580,604052]

estimated-cyto:21B7-21C7: 05/19/2001

Plate: SD.195 row: A column: 6

High quality sequence stop: 614.

FEATURES

source

1..637

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="SD19506"

/clone\_lib="SD Drosophila melanogaster Schneider L2 cell  
culture pot2"

/lab\_host="DH5-alpha"

/note="Vector: pot2; Site\_1: EcoRI; Site\_2: XhoI; Sized

fractionated cDNAs were directly ligated into pot2.

Plasmid cDNA library."

BASE COUNT 164 a 164 c 194 g 115 t

ORIGIN

Query Match 38.4%; Score 597.8; DB 13; Length 637;

Best Local Similarity 99.2%; Pred. No. 1.7e-87;

Matches 612; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 7 CAAAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCACATACGATACAGGCGGC 66  
DB 21 CAGAAAATGCAGGAGACGGACACGGAGCAAGAGGCCACGCACATACGATACAGGCGGC 80

QY 67 CTCGTTT---ACACGGGCGAGGAAGCCGCCGCCCAATCGACCTGCCGCGAGGAATCACT 123  
DB 81 CTCGTTTCGGACACACGGGCGAGGAAGCCGCCGCCCAATCGACCTGCCGCGAGGAATCACT 140

QY 124 ACCCAGCAATTGGGACTGATTTGCAACCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 183  
DB 141 ACCCAGCAATTGGGACTGATTTGCAACCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 200

QY 184 TTGTTTTTCGTTGGCGAGGATGAGATCAAGAAGAGCTGTGATCATCGAGGACACCTTGGCG 243  
DB 201 TTGTTTTTCGTTGGCGAGGATGAGATCAAGAAGAGCTGTGATCATCGAGGACACCTTGGCG 260

QY 244 TCAGTGGACACCGAAAAACGTGATTCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTG 303  
DB 261 TCAGTGGACACCGAAAAACGTGATTCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTG 320

QY 304 CGCCCAGTGACAAGATGACGAGTTCCATGCCGGGACACGCCGAGGCTGTGTTTCGCTG 363

Db 321 CGCCAGTGACAAGATGCACGAGTTCATGCGCGGACACGCCGAGGCTGTGGTTTCGCTG 380

QY 364 AATTTCAGCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCATTG 423

Db 381 AATTTCAGCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCATTG 440

QY 424 TGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGTGCATTAAGCACTGGTCTG 483

Db 441 TGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGTGCATTAAGCACTGGTCTG 500

QY 484 TGCGTATCCTGGCTCCGGATGGCAACCGTTGGCCACCGGTTGCCAAGCGGGCTCTATA 543

Db 501 TGCATATCCTGGCTCCGGATGGCAACCGTTGGCCACCGGTTGCCAAGCGGGCTCTATA 560

QY 544 ATCATCTGGACCCGGAGACGGGTACGAGAGGGGGCGACCCCTTGAGTGGGCACAAGAAA 603

Db 561 ATCATCTGGACCCGGAGACGGGTACGAGAGGGGGCGACCATTGAGTGGGCACAAGAAA 620

QY 604 CACATCAACTGCCTCGC 620

Db 621 CACATCAACTGCCTCGC 637

RESULT 3

BF505998 725 bp mRNA linear EST 19-APR-2001

LOCUS AT08344.5prime AT Drosophila melanogaster adult testes pOTB7

DEFINITION Drosophila melanogaster cDNA clone AT08344 5 similar to Nle: FBan0002863 'signal transduction' located on: 2L 21C6-21C6;; 04/07/2001, mRNA sequence.

ACCESSION BF505998

VERSION BF505998.2 GI:13688580

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 725)

AUTHORS Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan,D., Friese,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.

TITLE BDGP/HMI AT Drosophila EST Project

JOURNAL Unpublished (2000)

COMMENT On Dec 6, 2000 this sequence version replaced gi:11589299. Contact: Stapleton, M. BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, [est@fruitfly.berkeley.edu](mailto:est@fruitfly.berkeley.edu)

hit genomic AE003589: arm:2L [301580..604052]

estimated-cyto:21B7-21C7: 04/07/2001

Plate: Ar.83 row: D column: 8

High quality sequence stop: 661.

Location/Qualifiers

1..725

/organism="Drosophila melanogaster"

/db\_xref="taxon:7227"

/clone="AT08344"

/clone\_lib="AT Drosophila melanogaster adult testes pOTB7"

/sex="male"

/dev\_stage="0-3 day old Ore-R males"

/lab\_host="Plates Ar.10-AT.120: DH5-alpha. Plates AT.121-AT.319: DH5-alpha Tona"

/note="Organ: ADULT testes; Vector: pOTB7; Site.1: EcoRI; Site.2: XhoI; The mRNA for the testis library was made from testes and seminal vesicles hand dissected from 0-3 day old Ore-R males. RNA kindly provided by the lab of Margaret Fuller. Sized fractionated cDNAs were directly

FEATURES

source

BASE COUNT 187 a 182 c 215 g 141 t

ORIGIN

Query Match 38.1%; Score 591.8; DB 12; Length 725;

Best Local Similarity 99.2%; Pred. No. 1.5e-86;

Matches 606; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 7 CAAAAATGCAGGAGACGGACACCGGAGCAAGAGGCCACCCACATACAGGCGCGC 66

Db 115 CAGAAAATGCAGGAGACGGACACCGGAGCAAGAGGCCACCCACATACAGGCGCGC 174

QY 67 CTCGTTT---ACACGGGCGAGGAAGCGCGCCCGCCCAATCGACCTGCCGGCAGGAATCACT 123

Db 175 CTCGTTTCGGACACCGGCGAGGAAGCGCGCCCGCCCAATCGACCTGCCGGCAGGAATCACT 234

QY 124 ACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 183

Db 235 ACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGCCACTCCATAT 294

QY 184 TTGTTTTTCGTGGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 243

Db 295 TTGTTTTTCGTGGGCGAGGATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCG 354

QY 244 TCAGTGGACACCCGAAACGTCGATCGATATGTGTATCAGCCACAGGCGGTTTTCAAAGTG 303

Db 355 TCAGTGGACACCCGAAACGTCGATCGATATGTGTATCAGCCACAGGCGGTTTTCAAAGTG 414

QY 304 CGCCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGGTTTCGGCTG 363

Db 415 CGCCCAGTGACAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGGTTTCGGCTG 474

QY 364 AATTTTCAGCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCATTG 423

Db 475 AATTTTCAGCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCATTG 534

QY 424 TGGGATCTTAACACAGACACCCGCACTTCACCTGCACAGGTCATAAGCAGTGGGTTCTG 483

Db 535 TGGGATCTTAACACAGAGACACCCGCACTTCACCTGCACAGGTCATAAGCAGTGGGTTCTG 594

QY 484 TCGGTATCCTGGGCTCCGGATGGCAACCGTTGGCCAGCGGTTGCAAAGCGGGCTCTATA 543

Db 595 TCGGTATCCTGGGCTCCGGATGGCAACCGTTGGCCAGCGGTTGCAAAGCGGGCTCTATA 654

QY 544 ATCATCTGGACCCCGAGACGGGTACGAGAAGGGGGCGACCCCTTGAGTGGGCACAAGAAA 603

Db 655 ATCATCTGGACCCCGAGACGGGTACGAGAAGGGGGCGACCATTGAGTGGGCACAAGAAA 714

QY 604 CACATCAACTG 614

Db 715 CACATCAACTG 725

RESULT 4

BI632742 622 bp mRNA linear EST 10-SEP-2001

LOCUS SD26605.5prime SD Drosophila melanogaster Schneider L2 cell culture

DEFINITION pOT2 Drosophila melanogaster cDNA clone SD26605 5 similar to Nle: FBan0002863 GO:[signal transduction (GO:0004871)] located on: 2L 21C6-21C6;; 05/23/2001, mRNA sequence.

ACCESSION BI632742

VERSION BI632742.1 GI:15534952

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 622)

AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.

TITLE BDGP/HMI Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu  
hit genomic AE003589: arm:2L [301580,604052]  
estimated-cyto:21B7-21C7: 05/23/2001  
Plate: SD.266 row: A column: 5  
High quality sequence stop: 499.  
Location/Qualifiers  
1. .622  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="SD26605"  
/clone\_lib="SD Drosophila melanogaster Schneider L2 cell  
culture pOT2"  
/lab\_host="DH5-alpha"  
/note="Vector: pOT2; Site\_1: EcoRI; Site\_2: XhoI; Sized  
fractionated cDNAs were directly ligated into pOT2.  
Plasmid cDNA library."  
BASE COUNT 162 a 159 c 188 g 113 t  
ORIGIN  
Query Match 37.2%; Score 578.2; DB 13; Length 622;  
Best Local Similarity 98.9%; Pred. No. 2.6e-84;  
Matches 604; Conservative 0; Mismatches 3; Indels 4; Gaps 2;  
QY 7 CAAAAATGCAGGACGGACACGGAGCAAGAGCCACGCCACATACGATACGAGCGCGC 66  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
13 CAGAAATGCAGGACGGACACGGAGCAAGAGCCACGCCACATACGATACGAGCGCGC 72  
QY 67 CTCGTTT---ACACGGCGGAGGAAGCCGCCCAATCGACTGCCGGCAGGAATCACT 123  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
73 CTCGTTTCGGACACGGCGGAGGAAGCCGCCCAATCGACTGCCGGCAGGAATCACT 132  
QY 124 ACCCAGCAATTGGGACTGATTGCAACGGCTGCTGAAACGAGGAAGCCACTCCATAT 183  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
133 ACCCAGCAATTGGGACTGATTGCAACGGCTGCTGAAACGAGGAAGCCACTCCATAT 192  
QY 184 TTGTTTTCGTGGCGGAGTGAATCAAGAGAGCCTGGAGGACACGTTGGACTTGGCG 243  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
193 TTGTTTTCGTGGCGGAGTGAATCAAGAGAGCCTGGAGGACACGTTGGACTTGGCG 252  
QY 244 TCAGTGGACACCGAAAACGTGATCGATATTGTTATCAGCCACAGGGCGTTTCAAAGTG 303  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
253 TCAGTGGACACCGAAAACGTGATCGATATTGTTATCAGCCACAGGGCGTTTCAAAGTG 312  
QY 304 CGCCAGTGACAAGATGCACGAGTTCATGCGGGGACACGCCGAGGCTGTGTTTCGCTG 363  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
313 CGCCAGTGACAAGATGCACGAGTTCATGCGGGGACACGCCGAGGCTGTGTTTCGCTG 372  
QY 364 AATTTCAGCCCGGATGGTCTCATCTGCCAGTGAAGTGGCGACACACAGTGGATTG 423  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
373 AATTTCAGCCCGGATGGTCTCATCTGCCAGTGAAGTGGCGACACACAGTGGATTG 432  
QY 424 TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCAGTGGTTCTG 483  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
433 TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCAGTGGTTCTG 492  
QY 484 TGGGATCTCTGGGCTCCGGATGGCAAACGGTTGGCCAGGGTTGCAACGGGGCTCTATA 543  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
493 TGGGATCTCTGGGCTCCGGATGGCAAACGGTTGGCCAGGGTTGCAACGGGGCTCTATA 552  
QY 544 ATCATCTGGGACCCGGAGACGGGTGAGCAAGAGGGGGCCACCCTTGAGTGGGCACAAGAA 603  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
553 ATCATCTGGGACCCGGAGAC-GGTCAGCAGAAAGGGCGGACCATTGAGTGGGCACAAGAA 611  
QY 604 CACATCAACTG 614  
Db |||||||||||  
612 CACATCAACTG 622

RESULT 5  
BI232464  
LOCUS  
DEFINITION  
BI232464  
RE28535.5prime RE Drosophila melanogaster normalized Embryo pFlc-1  
Drosophila melanogaster cDNA clone RE28535 5 similar to Nle:  
FBan0002863 'signal transduction' located on: 2L 21C6-21C6;:  
04/12/2001, mRNA sequence.  
ACCESSION  
BI232464  
VERSION  
BI232464.1 GI:14700036  
KEYWORDS  
EST.  
SOURCE  
fruit fly.  
ORGANISM  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 606)  
REFERENCE  
Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson  
,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George  
,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S.,  
Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,  
Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin  
,G.M.  
BDGP/HMI RE Drosophila EST Project  
Unpublished (2001)  
Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu  
hit genomic AE003589: arm:2L [301580,604052]  
estimated-cyto:21B7-21C7: 04/12/2001  
Plate: RE.285 row: C column: 11  
High quality sequence stop: 546.  
Location/Qualifiers  
1. .606  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone="RE28535"  
/clone\_lib="RE Drosophila melanogaster normalized Embryo  
pFlc-1"  
/sex="male and female"  
/dev\_stage="0-24 hours mixed stage embryonic"  
/lab\_host="DH5-alpha Tona"  
/note="Organ: embryo; Vector: pFlc1; Site\_1: XhoI; Site\_2:  
BamHI; Library was kindly generated by Piero Carninci at  
the RIKEN. The library was normalized and excised using  
Cre recombinase. Plasmid cDNA library."  
BASE COUNT 150 a 153 c 185 g 118 t  
ORIGIN  
Query Match 35.3%; Score 549.2; DB 13; Length 606;  
Best Local Similarity 98.9%; Pred. No. 1.3e-79;  
Matches 564; Conservative 0; Mismatches 3; Indels 3; Gaps 1;  
QY 7 CAAAAATGCAGGACGGACACGGAGCAAGAGCCACGCCACATACGATACGAGCGCGC 66  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
37 CAGAAATGCAGGACGGACACGGAGCAAGAGCCACGCCACATACGATACGAGCGCGC 96  
QY 67 CTCGTTT---ACACGGCGGAGGAAGCCGCCCAATCGACTGCCGGCAGGAATCACT 123  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
97 CTCGTTTCGGACACGGCGGAGGAAGCGGGCCGCCCAATCGACTGCCGGCAGGAATCACT 156  
QY 124 ACCCAGCAATTGGGACTGATTGCAACGGCTGCTGAAAACGAGGAAGCCACTCCATAT 183  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
157 ACCCAGCAATTGGGCTCTGATTGCAACGGCTGCTGAAAACGAGGAAGCCACTCCATAT 216  
QY 184 TTGTTTTCGTGGCGGAGGATGAGATCAAGAGAGCCTGGAGGACACGTTGGACTTGGCG 243  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
217 TTGTTTTCGTGGCGGAGGATGAGATCAAGAGAGCCTGGAGGACACGTTGGACTTGGCG 276  
QY 244 TCAGTGGACACCGAAAACGTTGATCGATATTGTTGATCAGCCAGCGGTTTCAAAGTG 303  
|||



Db	277	TCAGTGGACACCGAAACGTCATCGATATTGTGTATCAGCCACAGGCGGTTTCAAAGTG	336
QY	304	CGCCCAAGTACAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG	363
Db	337	CGCCCAAGTACAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGTGGTTTCGCTG	396
QY	364	AATTTCAGCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCAGATTG	423
Db	397	AATTTCAGCCCGGATGGTGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCAGATTG	456
QY	424	TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGTGCATAAGAGTGGTTCTG	483
Db	457	TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGTGCATAAGAGTGGTTCTG	516
QY	484	TGCGTATCCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAAGCGGGCTCTATA	543
Db	517	TGCGTATCCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAAGCGGGCTCTATA	576
QY	544	ATCATCTGGACCCGGAGACGGGTCAGCAG	573
Db	577	ATCATCTGGACCCGGAGACGGGTCAGCAG	606
RESULT 6			
BI639652			
LOCUS	BI639652	626 bp mRNA linear EST 10-SEP-2001	
DEFINITION	SD22291.5prime SD Drosophila melanogaster Schneider L2 cell culture		
	POT2 Drosophila melanogaster cDNA clone SD22291 5 similar to Nle:		
	FBan0002863 GO:[signal transduction (GO:0004871)] located on: 2L		
	21C6-21C6:: 05/19/2001, mRNA sequence.		
ACCESSION	BI639652		
VERSION	BI639652.1	GI:15541862	
KEYWORDS	EST.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
	Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 626)		
AUTHORS	Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,		
	Lewis,S. and Rubin,G.M.		
TITLE	BDGP/HMI Drosophila EST Project		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Stapleton, M.		
	BDGP		
	Lawrence Berkeley National Lab		
	One Cyclotron Rd, Berkeley, CA 94720, USA		
	Fax: 510 486 6798		
	Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu		
	hit genomic AE003589: arm:2L [301580,604052]		
	estimated-cyto:21B7-21C7: 05/19/2001		
	Plate: SD.222 row: H column: 7		
	High quality sequence stop: 544.		
FEATURES	location/Qualifiers		
source	1. .626		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone="SD22291"		
	/clone_lib="SD Drosophila melanogaster Schneider L2 cell		
	culture pot2"		
	/lab_host="DH5-alpha"		
	/note="Vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized		
	fractionated cDNAs were directly ligated into pot2.		
	Plasmid cDNA library."		
BASE COUNT	155 a 158 c 188 g 124 t	1	others
ORIGIN			
	Query Match 35.3%; Score 548.8; DB 13; Length 626;		
	Best Local Similarity 98.9%; Pred. No. 1.5e-79;		
	Matches 563; Conservative 0; Mismatches 3; Indels 3; Gaps 1;		
QY	7	CAAAAATGCAGGACGACACGGACAGAGCGCCACATACATACAGGCGCGC	66

Db	58	CAGAAAATGCAGGACGGACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCGC	117
QY	67	CTCGTTT---ACACGGCGGAGGAAGCCGGCCCCGCAATCGACCTGCCGGCAGGAATCACT	123
Db	118	CTCGTTTCGGACACGGCGGAGGAAGCCGGCCCCGCAATCGACCTGCCGGCAGGAATCACT	177
QY	124	ACCAGCAATTGGGACTGATTTGCAACGGCTGCTGAAAAACGAGGAAGCCACTCCATAT	183
Db	178	ACCAGCAATTGGGACTGATTTGCAACGGCTGCTGAAAAACGAGGAAGCCACTCCATAT	237
QY	184	TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTTGGACTTGGCG	243
Db	238	TTGTTTTTCGTGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACGTTGGACTTGGCG	297
QY	244	TCAGTGGACACCCGAAAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTG	303
Db	298	TCAGTGGACACCCGAAAAACGTGATCGATATTGTGTATCAGCCACAGCGGTTTCAAAGTG	357
QY	304	CGCCCAAGTACACAGATGCACGAGTTCATGCCGGGACACCGCGAGGCTGTGGTTTCGCTG	363
Db	358	CGCCCAAGTACACAGATGCACGAGTTCATGCCGGGACACCGCGAGGCTGTGGTTTCGCTG	417
QY	364	AATTTAGCCCCGGATGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGGGATTG	423
Db	418	AATTTAGCCCCGGATGCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGGGATTG	477
QY	424	TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTGATAAGCAGTGGTTCTG	483
Db	478	TGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTGATAAGCAGTGGTTCTG	537
QY	484	TGCGTATCCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAGCGGGCTCTATA	543
Db	538	TGCGTATCCTGGGCTCCGGATGGCAACGGTTGGCCAGCGGTTGCAAGCGGGCTCTATT	597
QY	544	ATCATCTGGGACCCGGAGACGGGTCAGCA	572
Db	598	ATCATCTGGGACCCNGAGACGGGTCAGCA	626
RESULT 7			
AI520108			
LOCUS	AI520108	527 bp mRNA linear EST 19-APR-2001	
DEFINITION	LD40082.5prime LD Drosophila melanogaster embryo pot2 Drosophila		
	melanogaster cDNA clone LD40082 5prime, mRNA sequence.		
ACCESSION	AI520108		
VERSION	AI520108.1	GI:4425962	
KEYWORDS	EST.		
SOURCE	fruit fly.		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
	Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 527)		
AUTHORS	Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,		
	Lewis,S. and Rubin,G.M.		
TITLE	BDGP/HMI Drosophila EST Project		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Stapleton, M.		
	BDGP		
	Lawrence Berkeley National Lab		
	One Cyclotron Rd, Berkeley, CA 94720, USA		
	Fax: 510 486 6798		
	Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu		
	hit genomic sequence AC004573		
	Plate: 400 row: G column: 10		
	High quality sequence stop: 506.		
FEATURES	location/Qualifiers		
source	1. .527		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/clone="LD40082"		
	/clone_lib="LD Drosophila melanogaster embryo pot2"		
	/sex="male and female"		







```

ACCESSION      21C6-21C6;; 05/18/2001, mRNA sequence.
VERSION        BI635228
KEYWORDS       BI635228.1 GI:15537438
SOURCE         EST.
ORGANISM       fruit fly.
               Drosophila melanogaster
               Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
               Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
               Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 510)
AUTHORS        Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
               Lewis,S. and Rubin,G.M.
TITLE          BDGP/HMI Drosophila EST Project
JOURNAL        Unpublished (2001)
COMMENT        Contact: Stapleton, M.
               BDGP
               Lawrence Berkeley National Lab
               One Cyclotron Rd, Berkeley, CA 94720, USA
               Fax: 510 486 6798
               Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
               hit genomic AE003589: arm:2L [301580,604052]
               estimated-cyto:21B7-21C7: 05/18/2001
               Plate: SD.165 row: B column: 3
               High quality sequence stop: 229.
               Location/Qualifiers
               1..510
               /organism="Drosophila melanogaster"
               /db_xref="taxon:7227"
               /clone="SD16515"
               /clone_lib="SD Drosophila melanogaster Schneider L2 cell
               culture pot2"
               /lab_host="DH5-alpha"
               /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
               fractionated cDNAs were directly ligated into pOT2.
               Plasmid cDNA library."
               Plasmid cDNA library."
BASE COUNT     131 a 130 C 151 g 98 t
ORIGIN
Query Match    29.3%; Score 455.8; DB 13; Length 510;
Best Local Similarity 97.9%; Pred. No. 1.8e-64;
Matches 473; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY 7 CAAAAAATGCAGGACGGACAGCGGACAGAGGCCACGCACATACGATACAGGCGGC 66
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
28 CAGAAATGCAGGACGGACACGGAGGACAGAGGCCACGCACATACGATACAGGCGGC 87
QY 67 CTCGTTT---ACACGGGGGAGGAAGCGCGCCGCCAATCGACCTGCCGAGGAATCACT 123
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
88 CTCGTTTCGGACACGGGGGAGGAAGCGCGCGCTCAATCGACCTGCCGAGGAATCACT 147
QY 124 ACCAGCAATTGGGACTGATTGCAACGGCGTCTGTAACAGAGGAGGAGGAGGAGGAGG 183
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
148 ACCAGCAATTGGGACTGATTGCAACGGCGTCTGTAACAGAGGAGGAGGAGGAGGAGG 207
QY 184 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCTGGAGGACACCGTTGGCGG 243
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
208 TTGTTTTTCGTGGCGGAGGATGAGATCAAGAAGAGCCTGGAGGACACCGTTGGCGG 267
QY 244 TCAGTGGACACCGAAAACGTGATCGATATTGTGTATCAGCCACAGGCGGTTTCAAAGTG 303
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
268 TCAGTGGACACCGAAAACGTGATCGATATTGTGTATCATCCACAGGCGGTTTCAAAGTG 327
QY 304 CGCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACCGCGAGGCTGTGGTTCGCTG 363
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
328 CGCCAGTGACAAGATGCACGAGTTCCATGCCGGGACACCGCGAGGCTGTGGCTTCGCTG 387
QY 364 AATTTCAGCCCGGATGGTCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCAGTTG 423
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
388 AATTTCAGCCCGGATGGTCTCATCTCGCCAGTGGAAAGTGGCGACACACAGTGCAGTTG 447
QY 424 TGGGATCTTAACACAGAGACACCGCATTACCTGCACAGGTCAATAGCAGTGGGTTCTG 483
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
448 TGGGATCTTAACACAGAGACACCGCATTACCTGCACAGGTCAATAGCAGTGGGTTCTG 507
```

```

QY 484 TGC 486
Db 508 TGC 510

RESULT 11
AA735870
LOCUS     AA735870               448 bp      mRNA      linear      EST 23-APR-2001
DEFINITION
GM10004.5prime GM Drosophila melanogaster ovary Bluescript
Drosophila melanogaster cDNA clone GM10004 5prime, mRNA sequence.
ACCESSION AA735870
VERSION   AA735870.1 GI:2761800
KEYWORDS  EST.
SOURCE    fruit fly.
ORGANISM  Drosophila melanogaster
           Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
           Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
           Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 448)
AUTHORS    Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
           Lewis,S. and Rubin,G.M.
TITLE      BDGP/HMI Drosophila EST Project
JOURNAL    Unpublished (2001)
COMMENT    Contact: Stapleton, M.
           BDGP
           Lawrence Berkeley National Lab
           One Cyclotron Rd, Berkeley, CA 94720, USA
           Fax: 510 486 6798
           Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
           Plate: 100 row: A column: 4
           High quality sequence stop: 409.
           Location/Qualifiers
           1..448
           /organism="Drosophila melanogaster"
           /db_xref="taxon:7227"
           /clone="GM10004"
           /clone_lib="GM Drosophila melanogaster ovary Bluescript"
           /sex="female"
           /dev_stage="newly eclosed females: germarium-stage 6"
           /lab_host="SOLR"
           /note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
           Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
           Synthesis kit. Oligo dT-primed and directionally cloned at
           EcoRI and XhoI in Bluescript SK(+/-)"
BASE COUNT 112 a 121 c 131 g 84 t
ORIGIN
Query Match    27.3%; Score 424; DB 9; Length 448;
Best Local Similarity 98.2%; Pred. No. 2.6e-59;
Matches 440; Conservative 0; Mismatches 5; Indels 3; Gaps 1;

QY 25 GACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCGCTCGTT---ACACGGGC 81
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 GACACGGAGCAAGAGGCCACGCCACATACGATACAGGCGCGCTCGTTTCGGACACGGGC 60
QY 82 GAGGAAGCCGGCCGCCAATCGACCTGCCGGCAGGAATCACTACCCAGCAATTGGGACTG 141
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 GAGGAAGCCGGCCGCCAATCGACCTGCCGGCAGGAATCACTACCCAGCAATTGGGACTG 120
QY 142 ATTTGCAACGGCGTCTGTAACAGGAGGAGCCACTCCTATATTGTTTTCGTGGCGGAG 201
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 ATTTGCAACGGCGTCTGTAACAGGAGGAGCCACTCCTATATTGTTTTCGTGGCGGAG 180
QY 202 GATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCGTCAGTGGACACCGAAAAC 261
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 GATGAGATCAAGAAGAGCCTGGAGGACACGTTGGACTTGGCGTCAGTGGACACCGAAAAC 240
QY 262 GTGATCGATATTGTGTATCAGCCACAGGCGGTTTCAAAGTGGCGCCAGTGCACAGATGC 321
Db 11 |||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 GTGATCGATATTGTGTATCAGCCACAGCGCTGTTTCAAAGTGGCGCCAGTGCACAGATGC 300
QY 322 ACAGATTCCATGCCGGGACACGCCGAGGCTGTGGTTTCGTTGAATTCAGCCCGGATGGT 381
```



Db	301	ACGAGTCCATGCCGGACACGCCGAGGCTGTGGTTTCGCTGAATTCAGCCCGGATGGT	360	
QY	382	GCTCACTCGCCAGTGGAAAGTGGCGACACCACAGTCGATTGTGGGATCTTAACACAGAG	441	
Db	361	GCTCACTCGCCAGTGGAAAGTGGCGACACCACAGTCGATTGTGGGATCTTAACACAGAG	420	
QY	442	ACACCGCACTTCACCTGCACAGGTGCATA	469	
Db	421	ACACCGCACTTCACCTGCACAGGTGCATA	448	
RESULT 12				
BM579586				
LOCUS	BM579586	688 bp	mRNA	linear
DEFINITION	17000687239580 A.Gam.ad.cdNA.blood1 Anopheles gambiae	EST 22-FEB-2002		
ACCESSION	19600449699487 5', mRNA sequence.			
VERSION	BM579586			
KEYWORDS	BM579586.1 GI:18868053			
SOURCE	EST.			
ORGANISM	African malaria mosquito.			
REFERENCE	Anopheles gambiae			
AUTHORS	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.			
TITLE	1 (bases 1 to 688)			
JOURNAL	Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.			
COMMENT	Celera Anopheles gambiae EST project Unpublished (2002) Contact: Holt R.A. Celera Genomics 45 W. Gude Dr., Rockville, MD 20850, USA Tel: 2404533151 Fax: 2404534580 Email: HoltRA@celera.com Plate: NU01004AB3 row: D column: 01 Seq primer: M13 Reverse.			
FEATURES	source	Location/Qualifiers		
		1..688		
		/organism="Anopheles gambiae"		
		/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"		
		/db_xref="taxon:7165"		
		/clone="19600449699487"		
		/clone_lib="A.Gam.ad.cdNA.blood1"		
		/dev_stage="Adult"		
		/lab_host="DH10b"		
		/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'.		
		Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"		
BASE COUNT	136 a	178 c	240 g	134 t
ORIGIN				
	Query Match	22.3%	Score 346;	DB 13; Length 688;
	Best Local Similarity	71.0%;	Pred. No. 9e-47;	
	Matches 473;	Conservative	0; Mismatches 190;	Indels 3; Gaps 1;
QY	793	TCCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTGTGCCGGAGCTTC	852	
Db	14	TCCCGCGATCGGACCGTCAAGATGTGGCGAGCGGAGCGGTACTGTGC AAAACGTTTC	73	
QY	853	TCTGGCCAAAGCTCACTGGGTAACAACATTCGCGTGAGCACCGGATACGTCCTGCGCACT	912	
Db	74	ACGGGTCACGGCACTGGGTGAATAATTGGCGCTCAACACGGACTACGTGCTGCGGACG	133	
QY	913	GGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCCAC--CTCAGTTTGAGCACTGAG	969	
Db	134	GGCCCGTTTACCCCGTGATGGACAAATTTAAGATGACGGGGACGGCGGGATAGGAC	193	

QY	970	GAATTGCAGGAATCTGCCTTGAAGCGCTACCAGGCCGCTGCTCCCTGACGAGGTGGAGTCG	1029	
Db	194	GCCCTCAAGCAGCTGGCGCTCGAGCGGTACGAGCAGGTGTCCCGGACGGGTGGAATCG	253	
QY	1030	CTGGTTTCCTGTTCCGATGACAAACACCCTCTATCTGTGCGGGAACAACAGCAAGTGC	1089	
Db	254	TTCGTCTCGTGTTCGACGACTTTACGCTGTATCTGTGCGCGCGGAACACAGAAACAGTTC	313	
QY	1090	GTTGAGCGCATGACAGGGCACCAAGACGTGGTCAACGATGTGAAATATTGCCCGGATGTA	1149	
Db	314	GTGACGCGCATGACCGGCCACCAAGATGTGGTGAACGATGTGAAGTACTCGCCGACGTG	373	
QY	1150	AAGCTAATTGCGTCTGCTTCATTTGACAAAGTCAGTCCGCTGTGGCGAGCAGCGATGGT	1209	
Db	374	AAGTTTATTGCGTCCGCTTCGTTTCGACAAATCGATCCGGCTGTGGCGGCGCGGATGGG	433	
QY	1210	CAGTACATGCCACACCTTCCGGGGTTCATGTGCAGGCTGTTTACACGCTTGCCTGGTCCGCG	1269	
Db	434	CGGTTTCATTTGTGCTTGCCTTGCCTGCGGGTTCACGTGCTGCGCGTCTACACGGTGAGTGGT	493	
QY	1270	GACTCCCGCTGATTGTTTCCGGGCGAGCAAAAGACTCAACTTCAAAAAGTATGGAGTGTGCAG	1329	
Db	494	GATTCGCGGTAGTGTCTTAGCGGCGAGCAAGGACTCAACCTCAAGGTGTGGAGTGTGAAG	553	
QY	1330	ACGAAGAACTGGCACAGGAGTGCCTTGGACATGCGGATGAGGTGTTTCGAGTGGACTGG	1389	
Db	554	GAGCGCAAGCTGGTGCAGGAGTGCCTGCGGACACGCGGACGAGTGTACGCGTGGATTGG	613	
QY	1390	GCGCCCGATGGCTCTAGAGTTCCCTCTGTGGCAAGGACAAAGTTATAAGCTATGGCT	1449	
Db	614	GCACCGGACGGTCCCGCTGCTTCCGTTGGCAAGAGTGTATAAACTGTGGGCA	673	
QY	1450	TATTAA	1455	
Db	674	TATTAA	679	
RESULT 13				
AL520339				
LOCUS	AL520339	975 bp	mRNA	linear
DEFINITION	AL520339 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB006YB09 5 prime, mRNA sequence.			
ACCESSION	AL520339			
VERSION	AL520339.1	GI:12783832		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 975)			
TITLE	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.			
JOURNAL	Full-length cDNA libraries and normalization			
COMMENT	Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.			
FEATURES	source	Location/Qualifiers		
		1..975		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/clone="CS0DB006YB09"		
		/clone_lib="LTI_NFL004_NBC2"		
		/sex="male"		
		/tissue_type="neuroblastoma cells"		
		/lab_host="DH10B"		
		/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-Oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center		

Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"									
BASE COUNT		195 a	293 c	291 g	184 t	12 others			
ORIGIN									
Query Match		21.3%;	Score 331.6;	DB 9;	Length 975;				
Best Local Similarity		60.2%;	Pred. No. 1.7e-44;						
Matches 532;		Conservative	9;	Mismatches 343;	Indels	0;	Gaps	0;	
QY	54	GATACAGCGCGCCTCGTTTACACGGGCGAGGAAGCGCGGCCCGCAATCGACCTGCCGCG	113						
Db	52	GCTAGTGCAGTTCACAGATGAGGGCGGCAGCTGCTGGGTTCOCGCTTCGACGTGCCCGT	111						
QY	114	AGGAATCACTACCCAGCAATTGGGACTGATTTGCAACGCGCTGCTGAAAAACGAGGAAGC	173						
Db	112	GGACATCACCCCGGACAGGCTGCAGCTCGTGTGCAACGCGCTACTGGCCCCAGGAGGATCC	171						
QY	174	CACTCCATATTGTTTTTCGTGGGCGAGGATGAGATCAAGAAGAGCCCTGGAGGACACAGTT	233						
Db	172	CCTGCCACTGGCTTCTTTKTCACAGATGCTGAGATCGTCTCCTCACTGGGGGAARACGTT	231						
QY	234	GGACTTGGCGTCAGTGGACACCCGAAAACGTGATCGATATTGTGTATCAGCCACAGGCGGT	293						
Db	232	GGAGTCCCAGGCAGTGGAGACARARAAGGTCTTARACATCATCTACCAGCCACAGGCTAT	291						
QY	294	TTTCAAAGTGCGCCAGTGACAAAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGT	353						
Db	292	CTTCARAGTCCGGGCTGTGACTCGCTGCACCAGCTCCTTGGAGGTCACAGTGAGGCAGT	351						
QY	354	GGTTTCGCTGAATTTACGCCCGGATGGTGTCTCATCTGCCAGTGGAAAGTGGGACACACC	413						
Db	352	CATTTCTGTGGCCTTCARCCCTACGGGAAARTACCTGGCCAGTGGCTCTGGAGACACCAC	411						
QY	414	AGTGGGATTGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGTTCATAAGCA	473						
Db	412	CGTGGCCTTCTGGGATCTCAGCACAGAGACACCAATTTTACATGCAAGGGACACARACA	471						
QY	474	GTGGGTTCTGTGGGTATCTTGGGTCGGGATGGCAACGGTTGGCCAGCGGTTGCAAGC	533						
Db	472	CTGGGTCCTTAGTATATCCTGGTCTCCAGATGGCAAGAAGCTGGCCTCAGGCTGCAAGAA	531						
QY	534	GGGCTCTATAATCATCTGGGACCCGGAGACGGGTCAGCAGAGAAGGGCGACCCCTTGAGTGG	593						
Db	532	TGGCCAGATTCTCCTCTGGGACCCCAAGCACAGGGAAGCAGGTGGGCAGGACCCTCGCTGG	591						
QY	594	GCACAAGAAACACATCAACTGCCTCGCCTGGGAACCGTATCATCTCGGATCCGGAGTGCAG	653						
Db	592	CCACAGCAAGTGGATCACAGGCTCAGCTGGGAGCCCTCCATCGGAACCCCTGAGTGCCG	651						
QY	654	GAAACTTGCTTCCGCCAGTGGAGACGGGACTGCCGGATTGGGACGTAAATTTGGGCCA	713						
Db	652	CTATGTGGCCAGCAGCTCCAAGGATGGCAGTGTGCGGATCTGGGACACAACACTGCAGGCCG	711						
QY	714	GTGCCCTTATGAACATTGCCGGACACACAAATGCTGTGACAGCAGTGAGATGGGGTGGAGC	773						
Db	712	CTGTGAGCGCATCCTCACCGGGCACACCCCACTCGGTCACTGCTCCGGTGGGAGGGGA	771						
QY	774	GGGCTTATTATACATCCTCCAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG	833						
Db	772	CGGGCTTCTCTACTCTGCTCCAGGACCGCACCATCAAAATCTGGAGAGCTCATGACGG	831						
QY	834	AATCTTGTCCGGACGTTCTCTGGCCAAGCTCACTGGGTAAACAACATTTGGGCTGAGCAC	893						
Db	832	TGTGCTGTGCCGGACTCTGCAAGGCCACGGCCACTGGGTGAACACAATGGCCCTCARCAC	891						
QY	894	CGATTACGTCCCTGCCGACTGGTCCATTCCATCCGGTGAAGGATC	937						
Db	892	TGACTATGCCCTGCGCACTGGGGCCTTTGAACATGCTGAGGCCCTC	935						

RESULT 14  
AL526575

LOCUS	AL526575	925 bp	mRNA	linear	EST 13-FEB-2001
DEFINITION	AL526575 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC014YN14 5 prime, mRNA sequence.				
ACCESSION	AL526575				
VERSION	AL526575.1	GI:12790068			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 925)				
TITLE	Li,W.B., Gruber,C., Jessee,J. and Polayes,D.				
JOURNAL	Full-length cDNA libraries and normalization				
COMMENT	Unpublished (2001)				
	Contact: Genoscope				
	Genoscope - Centre National de Sequencage				
	BP 191 91006 EVRY cedex - France				
	Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.				
FEATURES	Location/Qualifiers				
source	1..925				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="CS0DC014YN14"				
	/clone_lib="LTI_NFL003_NBC3"				
	/sex="male"				
	/tissue_type="neuroblastoma cells"				
	/lab_host="DH10B"				
	/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"				
BASE COUNT	183 a	278 c	290 g	173 t	1 others
ORIGIN					
Query Match	20.2%; Score 314; DB 9; Length 925;				
Best Local Similarity	60.7%; Pred. No. 1.2e-41;				
Matches	529; Conservative	1; Mismatches	341; Indels	1; Gaps	1;
QY	54	GATACAGCGCGCCTCGTTTACACGGGCGAGGAGCCGCCCAATCGACCTGCCGCG	113		
Db	54	GCTAGTGCAGTTCACAGGATGAGGCGGCGCAGCTGCTGGGTTCCCGTTCGACGTGCCCGT	113		
QY	114	AGGAATCACTACCCAGCAATTGGGACTGATTGCAACGCGCTGCTGAAAAACGAGGAGC	173		
Db	114	GGACATCACCCCGGACAGGCTGCAGCTCGTGTGCAACGCGCTACTGGCCCGAGGATCC	173		
QY	174	CACTCCATATTGTTTTCGTGGCGGAGGATGAGATCAAGAAGACCTGGAGGACACGTT	233		
Db	174	CCTGCCACTGGCTTTCTTTGTCCACGATGCTGAGATCGTCTCCTCACTGGGGAAGACGTT	233		
QY	234	GGACTTGGCGTCAGTGGACACCCGAAAACGTGATCGATATTGTGTATCAGCCACAGGCGGT	293		
Db	234	GGAGTCCCAGGCAGTGGAGACAGAGAAGGTCTTAGACATCATCTACCAGCCACAGGCTAT	293		
QY	294	TTTCAAAGTGCGCCAGTGACACAGATGCACGAGTTCATGCCGGGACACGCCGAGGCTGT	353		
Db	294	CTTCAGAGTCCGGGCTGTGACTCGCTGCACCAGCTCCTTGGAGGTCACAGTGAGGCAGT	353		
QY	354	GGTTTCGCTGAATTTACAGCCCGGATGGTGTCTATCTCGCCAGTGGGAAGTGGCGACACCC	413		
Db	354	CATTTCGTGGCCTTCAGCCCTACGGGAAAGTACCTGGCCAGTGGCTCTGGAGACACCCAC	413		
QY	414	AGTGGGATTGTGGGATCTTAACACAGAGACACCGCACTTCACCTGCACAGGTCATAAGCA	473		
Db	414	CGTGGCTTCTGGGATCTCAGCACAGAGACACACATTTACATGCAAGGGACACAGACA	473		
QY	474	GTGGGTTCTGTGGCGTATCCTGGGCTCCGGATGGCAAAACGGTTGGCCAGCGGTTGCAAAAGC	533		

```

||||| || | ||||||| |||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 474 CTGGTCTTAGTATATCTGGTCTCCAGATGGCAAGAAGCTGGCCTCAGGCTGCAAGAA 533

QY 534 GGGCTCTATAATCATCTGGGACCGGAGACGGGTACAGAGAAGGGGGACCCCTTGAGTGG 593
||| || ||||||||||| ||| ||||||| ||| ||||||| ||| ||| |||
Db 534 TGGCCAGATTCTCTCTGGGACCCCAAGCACAGAGGGAAGCAGGTGGGACGACCTCGCTGG 593

QY 594 GCACAAGAAACACATCAACTGCCTCGCTGGGAACCGTATCATCGCGATCCGGAGTGCAG 653
|||| || |||| |||| ||||||| ||| ||||| ||| ||||| |||
Db 594 CCACAGCAAGTGGATCACAGGCCCTGAGCTGGAGCCCTCCATGCCAACCCTGAGTGCCG 653

QY 654 GAAACTTGCTTCGCCCACTGGAGACGGGACTGCCGGATTGGGACGTAAATTTGGGCCCA 713
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 654 CTATGTGGCCAGCAGCTCCAAAGGATGGCAGTGTGCGGATCTGGGACACAACACTGCAGCCG 713

QY 714 GTGCCCTTATGAACATTGCCGGGACACACAAAATGCTGTGACAGCAGTGAGATGGGTGGAGC 773
|| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 714 CTGTAGCGCATCCTCACCGGGCACACCCAGTCGGTCACTGTCTCCGGTGGGAGGGGA 773

QY 774 GGGCCTTATTATACATCTCTCAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGG 833
|| ||| || | | ||||| || ||||| || ||||| || ||||| || |||||
Db 774 CGGGCTTCTACTCTGCTCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCTCATGACGG 833

QY 834 AATCTTGTCCGGACGTTCTCTGGCCAAAGCTCACTGGGTAAACAACATTCGCTGAGCAC 893
| ||||||||| | | | | | ||||||| ||||| ||||| ||||| |||||
Db 834 TGTGCTGTGCCGGAC-TGTGCAAGGCACGGCCACTGGGTGAACACCATGGCCCTCARCAC 892

QY 894 CGATTACGTCCTGCGCACTGGTCCATTCCATC 925
|| || ||||||||||| || |||
Db 893 TGACTATGCCCTGCGCACTGGGGCCTTTGAAC 924
```

```

RESULT 15
AL530778
LOCUS
DEFINITION AL530778 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DD008YA24 5
prime, mRNA sequence.
ACCESSION AL530778
VERSION AL530778.1 GI:12794271
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1114)
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..1114
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DD008YA24"
/clone_lib="LTI_NFL001_NBC4"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 233 a 320 c 351 g 180 t 30 others
ORIGIN
```

```

Query Match 19.4%; Score 301.4; DB 9; Length 1114;
Best Local Similarity 57.3%; Pred. No. 1.2e-39;
Matches 582; Conservative 19; Mismatches 406; Indels 8; Gaps 4;

QY 425 GGGATCTTAACACAGAGACACCGCACTTCACTGCACAGGTCAATAAGCAGTGGTTCGT 484
||||| | ||||||||||| ||||||| ||| : ||| | ||| |||
Db 2 GGGAKCGCAGACAGAGACACCACTTTCACAGGCRAGGCACACAGACATGGGGCCTTA 61
||||| | ||||||||||| ||||||| ||| : ||| | ||| |||

QY 485 GCGTATCTCGGCTCCGGATGGCAACGGTTGGCCAGCGGTGCAAAAGCGGCTC-TATA 543
| ||| : ||||||| || ||||| : ||||| ||| : ||| |
Db 62 GGAGATSCGGGCTCCAGAGGGCAAGARGCGGGCCTCAGGCKGCRAGARTGGCGCAGAGT 121

QY 544 ATCATCTGGGACCCGGAGACGGGTACAGAGAAGGGCGACCCCTTGAGTGGGCACAAGAAA 603
|| ||||||||| : || ||||| ||| | ||| ||| |||
Db 122 CTCGCTGGGACCCAAAGRCAGGGAAGCAGKGGGCAGGACCCTCGCKGGCCACAGCAAG 181

QY 604 CACATCAACTGCCTCGCCTGGGAACCGTATCATCGCGATCCGGAGTGCAGGAAAACCTTGCT 663
||||| : || ||||| || ||||| ||| ||||||| | |||
Db 182 TGGATCACAGSSTGAGCGGGAGCCCCCTCCATGCGAACCTGAGTGCSCGATGTGGCC 241

QY 664 TCCGCCAGTGGAGACGGGACTGCCGGATTGGGACGTAAATTTGGCCAGTGCCTTATG 723
| | ||||| ||||| ||||| : ||||| || |||||
Db 242 AGCAGCTCCAAGGATGGCAGTGTGCGGRKSRGGGACACAACTGCAGGCCGCTGTGAGCGC 301

QY 724 AACATTGCCGACACACAAATGCTGTGACACAGTGAAGTGGGGTGGAGCGGCGCTTATT 783
| | : ||||| ||||| | ||||| : | ||||||| | |||||
Db 302 ATCKKACCCGGGCACACCCAGTCGGTCACCKGTSTCCGGTGGGGAGGGGACGGGCTTCTC 361

QY 784 TATACATCCTCCAAAAGATCGCACAGTGAAGATGTGGCGAGCAGCTGATGGAATCTTGTGC 843
|| : ||||| || ||||| || || ||||| ||||| ||||| : |||||
Db 362 TACHCTGCCCTCCAGGACCGCACCATCAAAAGTCTGGAGAGCKATGACGGTGTGCKGTGC 421

QY 844 CGGACGTTCTCTGGCCCAAGCTCACTGGGTAAACAACATTTGGCTGAGCACCGGATTACGTC 903
||||| | ||||| ||||||| ||||||| ||||| ||| ||||| |||||
Db 422 CGGACTCTGCAAGGCCACCGGCACCTGGGTGAACACCATGCGCCCTCAGCACTGACTATGCC 481

QY 904 CTGCGCACTGTGTCCATTCCATCCGGTGAAGGATCGCTCCAAGAGCC---ACCTCAGTTTG 960
|||||||||| | || | | : ||| || || |||||
Db 482 CTGCGCACTGGGGCCTTTGAACCTGCKGAGGCGCTCAGTGAATCCCCAAGACCTCCAAGGA 541

QY 961 AGCACTGAGGAATTGCAGGAATCTGCCTTGAAGCGCTACAGGCCGCTGTCGCCCTGACGAG 1020
| |||| | |||| | |||| | |||| | |||| | |||| |
Db 542 TCCTTGCAGGAGTGAAGGAGAGGGGCTCTGAGCCCGATACAACCTCGTGGGGGCCAGGGT 601

QY 1021 GTGGAGTGGTGGTTTCTCTGTCGGATGACAAACCCCTCTATCTGTGG---CGGAACAAC 1077
||| ||||||| || ||||| ||||| ||||| ||||| |
Db 602 CCAGAGAGGCTGGTGTCTGGCTCCGACGACTCACCTTATTCCTGTGGTCCCGCAGCAGAG 661

QY 1078 CAGAACAAAGTGGTTGAGCGCATGACAGGGACCGGACCCAGAACGTTGGTCAACGATGTGAAATAT 1137
| |||| | ||||||| ||||||| ||||| | ||||| |||||
Db 662 GACAAAAGCCCTCTCACTCGGATGACAGGACACCAAGCTCTCATCAACCGGCTGCTCTTC 721

QY 1138 TCGCCGGATGTAAAGCTAATTGCGTCTGCITTCATTTGACAAGTCAGTGGCTGTGTGGCGA 1197
|| |||| | |||| | |||| | |||| | ||||| |||||
Db 722 TCTCCTGACTCCCGCATCGTGGCTAGTGCCTCTCTTTGACAAGTCCATCAAGCTGTGGGAT 781

QY 1198 GCCAGCGATGGTCACTAGTACATGGCCACCTTCCGGGGTCAATGTGCAGGCTGTTTACACGGTT 1257
| ||| || ||||||| || ||||| || ||||| || ||||| || |||||
Db 782 GGCAGGACGGGCAAGTACCTGGCTTCCCTACCGGGCCACGTGGCTGCCGTGTACCAGATT 841

QY 1258 GCCTGGTCCCGGACTCCCGCTTGATTGTTTCCGGCAGCAAGAGACTCAACTCTAAAAGTA 1317
|| |||| | ||| | | | || ||||||| ||| ||||| ||| |||||
Db 842 GCGTGGTCAAGTCAAGTCCGGTCTCTGGTCAAGCAGCAGCAGTGAAGGTG 901

QY 1318 TGGAGTGTGCAGACGAAGAACTGGCACAGGAGCTGCCCTGCACATGCGGATGAGGTGTTT 1377
||| |||| | ||||| ||||| ||||| ||||| ||||| |||||
Db 902 TGGGATGTGAAGGCCCCAGAAAGTGGCCATGGACCTGCCCGGCCACCGCGGATGAGGTATAT 961

QY 1378 GGAGTGGACTGGGCGCCCGATGGCTCTAGAGTTGCCCTCTGGTGGCAAGGACAAAG 1432
| || ||||| || ||||| | ||||| || ||||| ||||| |||||
Db 962 GCTGTTGACTGGAGTCCAGATGSGC-CAGAAATRGCAAGTGGTGGGAAGGACAATG 1015
```

Search completed: June 2, 2003, 04:48:17  
Job time : 2090 secs

---